

# Digital communication. Trends and good practices

Coordinators

Javier Guallar; Mari Váñez  
Anna Ventura-Cisquella

**CUVICOM**

Curación, Visualización y Visibilidad en la Comunicación Digital

**EPI**

This work is part of the Project "Parameters and strategies to increase the relevance of media and digital communication in society: curation, visualisation and visibility (CUVICOM)". Grant PID2021-123579OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU.



# CUVICOM

Curación, Visualización y Visibilidad en la Comunicación Digital

**Mari Vázquez**

Universitat de Barcelona, Spain  
[marivalez@ub.edu](mailto:marivalez@ub.edu)

**Javier Guallar**

Universitat de Barcelona, Spain  
[jguallar@ub.edu](mailto:jguallar@ub.edu)



**Ediciones Profesionales de la Información SL**

Mistral, 36 (Cortijo del Aire)

18220 Albolote (Granada), Spain

Tel.: +34-639 878 489/ 608 491 521

<https://www.scimagoepi.com>

© Cuvicom - Ediciones Profesionales de la Información SL

1<sup>o</sup> edition, August 2025

ISBN: 978-84-125757-6-7

<https://doi.org/10.3145/cuvicom.eng>

<https://www.scimagoepi.com/producto/digital-communication-trends-and-good-practices/>

#### **Recommended form of citation:**

Guallar, J., Vázquez, M., & Ventura-Cisquella, A. (Coords) (2025). *Digital communication. Trends and good practices*. Ediciones Profesionales de la Información. <https://doi.org/10.3145/cuvicom.eng>

Javier Guallar; Mari Vázquez; Ana Ventura-Cisquella

Work distributed under a licenseCC BY-NC-SA 4.0

<https://creativecommons.org/licenses/by-nc-sa/4.0>



Keys to digital communication in our time (Introduction).....	4
<i>Javier Guallar, Mari Vázquez &amp; Anna Ventura-Cisquella</i>	
Web visibility and journalism: SEO tips and best practices .....	7
<i>Carlos Lopezosa, Rubén Alcaraz-Martínez, Mari Vázquez &amp; Mario Pérez-Montoro</i>	
Google services for journalists and media: Recommendations for Google Discover and Google News .....	15
<i>Carlos Lopezosa, Douglas Cordeiro, Olaya López-Munuera &amp; Javier Guallar</i>	
Content curation in independent and media newsletters. Good practices and recommendations .....	25
<i>Pere Franch &amp; Javier Guallar</i>	
Content curation in social media for scientific journals. Trends and best practices .....	37
<i>Wileidys Artigas, Jesús Cascón-Katchadourian &amp; Javier Guallar</i>	
Curation of sources and equality narratives in the media agenda.....	49
<i>Anna Ventura-Cisquella &amp; Javier Guallar</i>	
SEOVisFrame: best practices and method for evaluating the SEO optimization of storytelling data visualisations .....	69
<i>Rubén Alcaraz-Martínez, Mari Vázquez, Raquel Escandel &amp; Diana Bajaña</i>	
Semantic content and information visualization: A proposal for a data-driven communication narrative .....	85
<i>Mario Pérez-Montoro</i>	
Mobilizing news media audiences. Models for the Activation of Relevance and Engagement in Audiences .....	104
<i>Llúcia Castells-Fos, Lluís Codina &amp; Carles Pont</i>	
Factors for enhancing visibility in digital repositories: Metadata quality, interoperability standards, persistent identifiers, and SEO-GEO optimization .....	119
<i>Danilo Reyes-Lillo, Cristòfol Rovira &amp; Alejandro Morales-Vargas</i>	
AI in Ibero-American Newsrooms: use cases and best practices .....	134
<i>Alexis Apablaza-Campos &amp; Jaime Andrés Wilches Tinjacá</i>	
AI and image banks: A research methodology .....	148
<i>Pere Freixa, Mar Redondo-Areolas, Lluís Codina &amp; Carlos Lopezosa</i>	
Critical thinking and artificial intelligence in academia: A qualitative matrix analysis procedure for evaluating AI systems.....	161
<i>Lluís Codina, Elisenda Aguilera-Cora, Carlos Lopezosa &amp; Pere Freixa</i>	
Web of Science Research Assistant: Functional analysis and usage recommendations.....	174
<i>Carlos Lopezosa, Elisenda Aguilera-Cora, Lluís Codina &amp; Juan-José Boté-Vericad</i>	

# Keys to digital communication in our time (Introduction)

**Javier Guallar**

Universitat de Barcelona, Spain

<https://orcid.org/0000-0002-8601-3990>

**Mari Vállez**

Universitat de Barcelona, Spain

<https://orcid.org/0000-0002-3284-2590>

**Anna Ventura-Cisquella**

Universitat de Barcelona, Spain

<https://orcid.org/0000-0002-3677-8055>

---

Guallar, J., Vállez, M., & Ventura-Cisquella, A. (2025). Keys to digital communication in our time (Introduction). En J. Guallar, M. Vállez, & A. Ventura-Cisquella (Coords). *Digital communication. Trends and good practices* (pp. 4-6). Ediciones Profesionales de la Información. <https://doi.org/10.3145/cuvicecom.00.eng>

---



This book, which we have titled *Digital communication. Trends and best practices*, aims to offer the interested public a series of concrete and practical contributions, the result of some of the academic research carried out over the last four years within the framework of the project “Parameters and strategies to increase the relevance of media and digital communication in society: Curation, visualization and visibility (CUVICOM).”

This work has been conceived to be of use and interest to a very diverse audience: professionals, academics, students, scholars, and anyone interested in the broad and dynamic world of digital communication.

Throughout its chapters, different specialists address topics ranging from search engine optimization to artificial intelligence, information visualization, and content curation, all guided by a practical perspective and intention.

What follows is a brief summary, offered as an introductory sample, of each chapter.

- **Web visibility and journalism: Tips and best SEO practices.** Explores the strategic integration of SEO in digital journalism, highlighting its role in enhancing the visibility, credibility, and economic sustainability of media outlets. It also offers practical recommendations that promote an ethical and quality-driven approach to content production.
- **Google services for journalists and media: Recommendations for Google Discover and Google News.** Analyzes Google News and Google Discover as key platforms for the visibility of digital media today. Their functionalities and implications are detailed, offering recommendations for content optimization.
- **Content curation in independent and media newsletters: Best practices and recommendations.** Addresses the current rise of newsletters as an effective communication product, emphasizing the importance of curation in connecting with audiences and providing specialized information. It also proposes recommendations for both media outlets and professionals.
- **Content curation in social media for scientific journals: Trends and best practices.** Focuses on how scientific journals use social media, with particular emphasis on content curation and the social network X, offering practical recommendations for journal editors and managers.
- **Curation of sources and equality narratives in the media agenda.** Analyzes the work of journalists specializing in equality in the Spanish digital press, with an emphasis on the curation of documentary sources and the topics they address, showing how they make visible inequalities, gender-based violence, feminist advances, and LGTBIQ+ rights.
- **SEOVisFrame: Best practices and method to evaluate SEO optimization of narrative data visualizations.** Presents the key positioning factors for narrative data visualizations in digital media, proposing the SEOVisFrame methodology to evaluate the SEO optimization of these informational products.
- **Semantic content and information visualization: A proposal for a data-based communication narrative.** Introduces the main communication contexts and semantic content that can be conveyed through graphics, offering a methodology for proper information visualization.

- **Mobilizing news media audiences: Models for the Activation of Relevance and Engagement in Audiences.** Offers a practical methodology for media organizations to assess and strengthen their relevance and engagement, crucial aspects of sustainability, broken down into three dimensions: reputation, visibility, and audience loyalty.
- **Factors for enhancing visibility in digital repositories: Metadata quality, interoperability standards, persistent identifiers, and SEO-GEO optimization.** Explores key strategies to improve the visibility of digital repositories, which are fundamental to increasing the reach and impact of academic content, such as metadata optimization, interoperability, and the adoption of persistent identifiers.
- **AI in Ibero-American newsrooms: Use cases and best practices.** Shows the integration of artificial intelligence into Ibero-American journalistic routines, highlighting its various uses such as virtual presenters, automated writing, audio conversion, assisted SEO, article summaries, and chatbots.
- **AI and image banks: A research methodology.** Provides a methodological framework for analyzing gender bias and sociocultural stereotypes in professional image banks. It also includes AI-generated images, revealing some stereotyped patterns and persistent clichés in visual content.
- **Critical thinking and Artificial Intelligence in academia: A qualitative Matrix Analysis Procedure for Evaluating AI Systems.** Proposes the Matrix Analysis Procedure for AI Systems (MASIA), a qualitative method for assessing AI performance in academia. This method focuses on narrative synthesis, source usage, and prompt formulation, fostering critical thinking and ethics in research.
- **Web of Science Research Assistant: Functional analysis and usage recommendations.** The final chapter, offers a functional analysis of Web of Science Research Assistant, a recent generative AI tool by Clarivate integrated into its well-known database that supports various tasks such as literature reviews and streamlining scientific discovery.

Taken as a whole, *Digital Communication. Trends and Best Practices* is a work of practical utility, which not only diagnoses the current state of digital communication but also offers a wide variety of tools, methodologies, best practices, and recommendations for professionals and academics. Both ethical and technical challenges of this digital era have been addressed, from the evolution of SEO or content curation in different domains to the need for the responsible integration of artificial intelligence, underscoring the importance of a commitment to informational quality. This book aims to serve as a resource for support, guidance, reference, or even inspiration for those who seek to stay up to date and avoid becoming disoriented in these times of intense digital transformation in the field of communication.

Javier Guallar, Mari Vázquez & Anna Ventura-Cisquella  
Barcelona, August 2025

# Web visibility and journalism: SEO tips and best practices

**Carlos Lopezosa**

Universitat de Barcelona, Spain

<https://orcid.org/0000-0001-8619-2194>

**Rubén Alcaraz-Martínez**

Universitat de Barcelona, Spain

<https://orcid.org/0000-0002-7185-0227>

**Mari Vállez**

Universitat de Barcelona, Spain

<https://orcid.org/0000-0002-3284-2590>

**Mario Pérez-Montoro**

Universitat de Barcelona, Spain

<https://orcid.org/0000-0003-2426-8119>

---

Lopezosa, C., Alcaraz-Martínez, R., Vállez, M., & Pérez-Montoro, M. (2025). Web visibility and journalism: SEO tips and best practices. In J. Guallar, M. Vállez, & A. Ventura-Cisquella (Coords). *Digital communication. Trends and good practices* (pp. 7-14). Ediciones Profesionales de la Información. <https://doi.org/10.3145/cuvicom.01.eng>

---

## Abstract

This chapter explores the strategic integration of Search Engine Optimization (SEO) in digital journalism, emphasizing its role in enhancing the visibility, credibility, and economic sustainability of media outlets. It begins by defining SEO and its evolution toward semantic SEO and Search Experience Optimization (SXO), stressing the balance between algorithmic optimization and journalistic integrity. The chapter outlines best practices for keyword research, headline optimization, semantic markup, and user experience improvements, providing actionable recommendations for journalists, editors, and technical teams. Finally, it discusses SEO as a critical component of sustainable business models in media, advocating for an ethical and quality-driven approach to digital content production.

## Keywords

SEO; Digital journalism; Web visibility; SXO; Keyword research; Semantic SEO; user experience; Media sustainability; Newsrooms; Content optimization.

---

## 1. Introduction

Search engine optimization (SEO) is a key factor for online media outlets (Codina *et al.*, 2016), as it determines how many potential users (readers) can access the news published on the internet (Dick, 2011). To ensure that news articles and reports reach the widest possible audience, journalists must apply good SEO practices—a set of strategies aimed at improving the ranking of content such as news, articles, reports, etc., in search engines like Google.

SEO in the journalistic field involves aspects such as researching relevant keywords and optimizing headlines (Lopezosa & Codina, 2018), making optimal use of tags and categories (Giomelakis & Veglis, 2015a), structuring texts clearly (Iglesias-García & Codina, 2016), and properly using technical elements such as meta descriptions and ALT tags for images (Giomelakis & Veglis, 2015b). In addition, it is essential to create original, well-documented, and high-quality content (Lopezosa *et al.*, 2019).

Implementing all these strategies not only increases visibility and organic traffic, but also helps strengthen the credibility of the media outlet and makes it easier for readers to find reliable and relevant information (Lopezosa & Codina, 2018).

This chapter aims to provide a practical guide to understanding how SEO can be ethically and effectively integrated into journalistic newsrooms.

## 2. Fundamentals of SEO in digital media

As previously mentioned, SEO refers to the set of strategies and techniques aimed at improving the visibility of a website in search engines, primarily Google (Giomelakis & Veglis, 2015a). In the context of journalism, this means maximizing the chances that a news article, report, or informational content appears among the top search results for a user's related query (Iglesias-García & Codina, 2016).

The importance of SEO for media outlets has two main components: not only does it allow a large portion of readers to access content through search queries (Asser, 2012), but it also enables media organizations to reach multiple target audiences, increase their influence as digital media, and better monetize their content—whether through advertising or subscription models.

One might think, then, that media organizations are forced to think more about search engines than readers, but nothing could be further from the truth. It is important to emphasize that integrating SEO strategies into journalistic production routines does not mean subordinating informational quality to the demands of Google's algorithm. Rather, it involves adapting content to make it more accessible, discoverable, and visible (Smyrnaio & Rebillard, 2009).

Clearly, the rise of search engines as the main gateway to news has forced media outlets to reconfigure their routines, but as we will see, what ultimately matters is the quality of the content and keeping citizens at the centre (Ferran-Ferrer et al., 2013).

In addition, SEO has evolved significantly since its inception. The way it was done in the past is very different from how it is done today—and it will change in the future. In fact, in its early phase, known as traditional SEO, optimization was based on factors such as keyword density, HTML structure, or the use of links. However, this highly technical and rigid approach soon showed its limitations as Google's algorithms began to prioritize content quality and user experience (Smyrnaio, 2015).

This led to the development of approaches such as semantic SEO and SXO (Search Experience Optimization). Semantic SEO emphasizes the meaning of content, using tools such as structured data, semantic tags (like those from Schema.org), and writing that aligns with natural language (Lopezosa et al., 2018). At the same time, SXO focuses on user experience, highlighting the need for content to be useful, easy to navigate, accessible, and capable of fostering interaction between the website and the user (Alcaraz-Martínez, 2024).

Today, both approaches—semantic and experiential—are increasingly integrated and complement each other within the visibility strategies of digital media (Giomelakis, 2023).

Google's Search Quality Rating Guidelines, first published in 2013 and continuously updated (Google, 2025), offer valuable insight into the criteria used by the search engine to evaluate website quality. Among the most significant concepts outlined in these guidelines are the EEAT principles—Experience, Expertise, Authoritativeness, and Trustworthiness (Alcaraz-Martínez, 2024). These principles address various dimensions related to content, authorship, and the overall credibility of a website, with the aim of determining whether the information presented is reliable—specifically, whether it is produced by individuals with relevant knowledge and whether it reflects direct or professional experience.

Google places particular emphasis on applying these principles in sensitive domains such as health, finance, law, and journalism, where content quality, accuracy, and editorial responsibility are critical to preventing misinformation and safeguarding users. Experience assesses whether the author possesses first-hand knowledge of the subject matter. Expertise refers to the author's education, competence, or specialized knowledge in the field. Authoritativeness evaluates whether the author, page, or website is recognized as a reputable source within its domain. Finally, Trustworthiness measures whether the content is secure, accurate, truthful, and transparent.

### 3. SEO Best practices for journalists and newsrooms

Below are some of the most common SEO practices used in news media, specifically focusing on keyword selection and optimized writing, dual headlines, semantic markup, and the optimization of design, navigation, and website architecture.

To a large extent, any search engine optimization strategy begins with the identification of relevant keywords (Smyrnaio & Sire, 2014). In the journalistic sector, these keywords should accurately reflect the core topic of the news story and match the search terms users commonly employ to find information. To carry out keyword research, journalists can rely on third-party tools such as Google Trends, SEMrush, Ahrefs, Majestic, Sistrix, or even Google's own auto-complete and suggested search terms.

Once the primary and secondary keywords have been identified, they can typically be integrated into the headline (H1 tag), the first paragraph, subheadings (H2, H3), the body of the text, the URL, and the meta descriptions.

This integration should be done naturally—without forcing the writing or compromising the content's quality. Therefore, narrative fluency and journalistic rigor remain priorities above any SEO tactic.

Another widely used SEO strategy in newsrooms is the practice of dual headlines (Asser, 2012). This involves using a short, direct headline on the media outlet's homepage, while applying a second, more SEO-optimized headline within the article itself. This practice, first adopted by the BBC and later standardized, allows media outlets to combine editorial impact with web visibility (Smyrnaio, 2015).

Beyond the headline, it is also recommended to structure content strategically. This usually involves including hierarchical subheadings using H2 and H3 tags to guide the reader, writing short paragraphs in clear and simple language, using lists, and incorporating internal and external links that provide added value to the reader.

Additionally, it is essential to apply semantic markup correctly, using technologies such as Schema.org. For news media, it is recommended to at least implement the NewsArticle schema. This can be added directly to the HTML code or through specialized plugins within the content management systems (CMS) used by each media outlet. The result is the appearance of rich snippets in search results—enhanced listings that may include news headlines, ratings, images, dates, authors, categories, knowledge graphs, etc.—which both increase click-through rates and enhance the perception of the outlet as a trusted source (Giomelakis, 2023).

As already mentioned, SEO has moved beyond simply ranking content through keywords or backlinks. In recent years, search engines—especially Google—have begun to consider user behaviour as a quality signal. This has led to the emergence of Search Experience Optimization (SXO) (Alcaraz-Martínez, 2024), which combines SEO and usability to ensure that content is both visible and meets the reader's expectations.

In journalism, this means that a news article must not only be well-ranked, but also deliver a satisfying reading experience through clear, useful content and optimized navigation. If users quickly abandon the page, do not interact, or fail to navigate to other articles, Google interprets this as a negative signal, which can harm the website's ranking.

Within SXO (Search Experience Optimization), website architecture plays a key role, as its main objective is to encourage reader retention and enhance their experience while navigating the site. In general, some key recommendations to improve user experience in this regard include:

- Responsive website design, meaning the digital news platform must adapt seamlessly to mobile devices and tablets.
- Optimized loading speed, as slow pages reduce the time users stay on the site.
- Clear hierarchical structure, with well-organized menus, categories, and tags.
- Breadcrumb navigation, to guide users effectively and help them understand their location within the site.
- Internal linking, encouraging readers to explore related, valuable content.

All these strategies are part of a semantic architecture, which not only improves the user experience but also helps search engines interpret and rank a website's content more effectively.

In addition, search engines use multiple interaction metrics to determine whether a piece of content satisfies the user and meets quality standards. Some of the most relevant indicators include:

- Bounce rate: the percentage of users who visit a website and leave without interacting.
- Dwell time: the average duration of a visit.
- Pages per session: how many pieces of content a user consumes during a single visit.
- CTR (Click-Through Rate): the percentage of clicks a result receives in relation to the number of impressions.
- Pogo-sticking: clicking on a result and immediately returning to the search results page to choose another link, which signals user dissatisfaction.
- Core Web Vitals: a set of specific metrics defined by Google to evaluate key aspects of user experience, related to speed, interactivity, and visual stability.
- Other technical signals: mobile compatibility, use of HTTPS, etc.

These indicators help understand and improve the user experience, and therefore also impact SEO and SXO. In the journalistic sector, this ultimately translates into well-written content that provides added value, includes engaging visuals, useful links, and a clear navigation structure.

## 4. Business models in digital media and sustainability through SEO

Beyond its importance as a visibility strategy, search engine optimization (SEO) is a key tool for the economic sustainability of digital media. In other words, strong search engine rankings not only attract more readers but also improve a media outlet's ability to monetize web traffic through advertising, subscriptions, sponsorships, and other hybrid models.

Organic traffic resulting from well-implemented SEO (focused on the user) is especially valuable because it is generally stable, targeted, and free in terms of acquisition costs.

However, media sustainability cannot be achieved through technical or commercial means alone. It must be grounded in a commitment to quality, independence, and journalism's social role. In this regard, SEO can be a strategic ally—as long as it is used ethically. That is, search



optimization strategies must respect content integrity, avoid information manipulation, and help quality journalism reach the widest possible audience.

The challenge, therefore, lies in integrating SEO into a responsible editorial culture, where metrics do not replace journalistic values, but rather serve as a tool to support high-quality information.

## 5. SEO Recommendations for media outlets

Incorporating SEO into journalistic newsrooms does not require a radical transformation, but rather a smart and gradual integration. Below, we present a set of practical recommendations that can serve as a guide for building an SEO culture within media organizations.

For reporters and journalists:

- Use clear, informative headlines with relevant and strategic keywords.
- Organize content using hierarchical subheadings, especially H2 and H3 tags.
- Include internal links to other content from the same media outlet to expand on the information being presented.
- Cite external sources using high-quality, relevant links.
- Add images with appropriate alternative text (alt attribute) and descriptive titles.
- Improve the perception of accuracy, truthfulness, and transparency of the content.

For editors and content managers:

- Implement an editorial policy that incorporates SEO criteria.
- Monitor publications in terms of length, frequency, and quality.
- Regularly review the headlines and meta descriptions of published articles.
- Apply dual headline techniques when appropriate.
- Showcase each writer's expertise and specialized knowledge through author pages with biographies and content.
- Train the entire team in SEO, usability, and user experience.

For technical and web teams:

- Ensure a clear and scalable website architecture.
- Optimize page loading speed and mobile/tablet responsiveness.
- Integrate semantic markup (Schema.org), at a minimum using the NewsArticle schema.
- Generate updated sitemaps (general and news-specific).
- Use SEO analysis tools (such as Google Search Console, Ahrefs, or SEMrush) to monitor the ranking and visibility of news content.

About the tools:

There are numerous tools—both free and paid—that support the practical application of SEO in media environments.



Type	Tools
For SEO auditing	Screaming Frog, SEMrush, Sistrix
For user behaviour analysis	Google Analytics, Hotjar, Matomo
For backlink and web authority analysis	Majestic, Ahrefs

It is worth noting that while having SEO knowledge within newsrooms is very important, the essential goal is to build an SEO mindset across the organization. To achieve this, key steps may include fostering collaboration between writers, editors, and SEO specialists, and helping the team understand that quality and journalistic ethics must always take precedence over rankings and clickbait.

Ultimately, a newsroom oriented toward implementing SEO strategies not only enhances the visibility of its content, but also helps improve the quality of its journalistic product, broadens its social impact, and strengthens its sustainability.

## 6. Conclusions

Throughout this chapter, we have aimed to provide a practical overview of the key concepts, techniques, and recommendations for applying SEO in the media sector.

As we have seen, integrating SEO ethically and professionally into journalistic newsrooms does not mean abandoning the principles of good journalism, but rather adapting them to the demands of the present. For example, an optimized headline is not incompatible with truthfulness, a well-constructed link does not diminish the quality of a report, and a user-centred architecture does not weaken the narrative. On the contrary, all these elements strengthen the final outcome of journalistic products.

We would like to conclude this chapter by emphasizing that SEO should not be seen as a threat, but rather as a tool in service of journalism. In fact, we believe that when applied with judgment, rigour, and professional sensitivity, quality content reaches more people, at the right time, and in the most accessible format.

Therefore, far from distorting the profession, SEO can help amplify its social impact—reinforcing journalism’s commitment to truth, its public value, and the citizen’s right to be informed.

## 7. Funding

This work is part of the Project “Parameters and strategies to increase the relevance of media and digital communication in society: curation, visualisation and visibility (CUVICOM)”. Grant PID2021-123579OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU.

## 8. References

Alcaraz-Martínez, R. (2024). Integración de la optimización de la experiencia de búsqueda (SXO), la usabilidad, la arquitectura de la información y la accesibilidad web: explorando un terreno común para mejorar la visibilidad en buscadores. *Ibersid: Revista de Sistemas de Información y Documentación*, 18(1), 37–53. <https://doi.org/10.54886/ibersid.v18i1.4966>

- Asser, M. (2012). Search engine optimisation in BBC News. <https://bbc.in/2J8TzIR>
- Codina, L., Iglesias-García, M., Pedraza, R., & García-Carretero, L. (2016). Visibilidad y posicionamiento web de informaciones periodísticas: El framework SEO-RCP. DigiDoc-UPF. <http://hdl.handle.net/10230/26040>
- Dick, M. (2011). Search engine optimization in UK news production. *Journalism Practice*, 5(4), 462-77. <https://doi.org/10.1080/17512786.2010.551020>
- Ferran-Ferrer, N., Minguillon, J., & Pérez-Montoro, M. (2013). Key factors in the transfer of information-related competencies between academic, workplace and daily life contexts. *Journal of The American Society for Information Science and Technology*, 64(6), 1112-1121. <https://doi.org/10.1002/asi.22817>
- Giomelakis, D., & Veglis, A. (2015a). Employing search engine optimization techniques in on-line news. *Studies in Media and Communication*, 3(1), 22-33. <https://doi.org/10.11114/smc.v3i1.683>
- Giomelakis, D., & Veglis, A. (2015b). Investigating search engine optimization factors in media websites, the case of Greece. *Digital Journalism*, 4(3), 379-400. <https://doi.org/10.1080/21670811.2015.1046992>
- Giomelakis, D. (2023). Semantic search engine optimization in the news media industry: Challenges and impact on media outlets and journalism practice in Greece. *Social Media+ Society*, 9(3). <https://doi.org/10.1177/20563051231195545>
- Google (2025). Search quality rating guidelines. <https://static.googleusercontent.com/media/guidelines.raterhub.com/es//searchqualityevaluatorguidelines.pdf>
- Iglesias-García, M., & Codina, L. (2016). Los cybermedios y la importancia estratégica del posicionamiento en buscadores (SEO). *Opción*, 32(9), 929-944. <https://bit.ly/2J8Skcu>
- Lopezosa, C., & Codina, L. (2018). *Análisis de posicionamiento en medios de comunicación con herramientas SEO: cobertura informativa de los premios Oscar 2017*. Universitat Pompeu Fabra. Departament de Comunicació. (Serie Editorial DigiDoc); (EPI; 2)
- Lopezosa C., Codina, L., & Caldera-Serrano, J. (2018) SEO semántico: framework ISS para la optimización de sitios intensivos en contenidos. *Cuadernos de documentación multimedia*, 29, 97-123. <https://doi.org/10.5209/CDMU.60607>
- Lopezosa, C., Codina, L., & Rovira, C. (2019). Visibilidad web de portales de televisión y radio en España: ¿qué medios llevan a cabo un mejor posicionamiento en buscadores. Serie DigiDoc-EPI, 4. <http://hdl.handle.net/10230/36234>
- Smyrnaio, N. (2015). Google and the algorithmic infomediatio of news. *Media Fields Journal*, 10, 1-10. <https://bit.ly/2XJIMN7>
- Smyrnaio, N., & Rebillard, F. (2009). L'actualité selon Google L'emprise du principal moteur de recherche sur l'information en ligne. *Communication & Langages*, 160, 95-109. <https://doi.org/10.4074/s0336150009002087>
- Smyrnaio, N., & Sire, G. (2014). The news according to Google how does algorithmic infomediatio frame the work of French journalists? In *JSS-ECREA 2014 Conference*, Thessaloniki. <https://bit.ly/33bA5tN>

# Google services for journalists and media: Recommendations for Google Discover and Google News

**Carlos Lopezosa**

Universitat de Barcelona, Spain

<https://orcid.org/0000-0001-8619-2194>

**Douglas F. Cordeiro**

Universidade Federal de Goiás, Brazil

<https://orcid.org/0000-0002-5187-0036>

**Olaya López-Munuera**

Universitat Oberta de Catalunya, Spain

<https://orcid.org/0009-0000-6132-5384>

**Javier Guallar**

Universitat de Barcelona, Spain

<https://orcid.org/0000-0002-8601-3990>

---

Lopezosa, C., Cordeiro, D. F., López-Munuera, O., & Guallar, J. (2025). Google services for journalists and media: Recommendations for Google Discover and Google News. In J. Guallar, M. Váñez, & A. Ventura-Cisquella (Coords). *Digital communication. Trends and good practices* (pp. 15-24). Ediciones Profesionales de la Información. <https://doi.org/10.3145/cuvicom.02.eng>

---

## Abstract

This chapter explores the role of Google News and Google Discover as key platforms for enhancing the visibility of digital media. Through a practical approach, it analyzes their functionalities, implications, and differences, providing specific recommendations for journalists and editors aiming to optimize content on these services. Google News functions as a news aggregator focused on recency and credibility, while Google Discover offers personalized content recommendations based on user behaviour. Best practice guidelines are proposed for both platforms, aiming to improve digital presence and face algorithmic challenges without compromising journalistic quality.

## Keywords

Google News; Google Discover; Digital journalism; SEO; Web visibility; Algorithms; Media; editorial strategies.

---

## 1. Introduction

The necessary adaptation of the journalistic sector to advances on the Internet has led, over the years, to significant transformations in the productive routines of news media (Salaverría & García-Avilés, 2008), especially regarding how news is disseminated through visibility in search engines (Iglesias-García & Codina, 2016) and content aggregation platforms (Lopezosa et al., 2019).

In this context, companies like Google —with specific services such as Google Discover and Google News— have become essential tools for media outlets to increase their visibility, broaden their reach, and consequently improve their monetization potential (Lopezosa et al., 2022; Cordeiro et al., 2024).

Google News is a content aggregation platform that collects and organizes news content from a wide variety of media sources, making it easier for readers to access current and reliable information (Guallar, 2015). Therefore, appearing on Google News represents not only an increase in readership for media organizations but also an improvement in their credibility (Lopezosa et al., 2024).

Google Discover, on the other hand, is a content recommendation platform that has become one of the main sources of traffic for many digital media outlets, especially on mobile devices (Lopezosa et al., 2022) and, since mid-2025, on desktop versions in some countries (González, 2025). Its ability to provide personalized suggestions and modify the feed in real time without requiring active search input from users influences editorial strategies, as it enables media outlets to reach new audiences (Lopezosa et al., 2023).

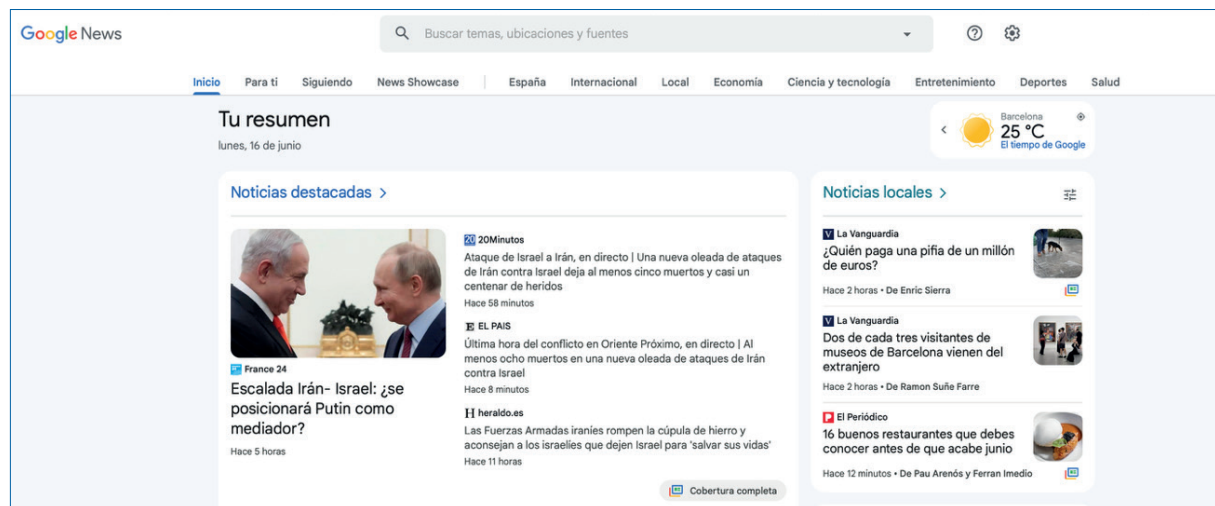
This chapter aims to analyze both Google services from a practical perspective, focusing on how content should be optimized so that media organizations can improve their presence on these two platforms.

## 2. What is Google News: How it works and what it means for the media

Google News is a content aggregator that collects headlines daily from hundreds of news sources and ranks and structures them based on each user's profile (reader). It is a free service and therefore does not include advertising (Lopezosa et al., 2024).

Image 1

Google News Front Page (06/16/2025).



In this way, Google News acts as a distributor of content from a wide range of information sources, which are characterized by being selected and accredited digital media outlets. Therefore, the main goal of Google News is to provide users with a broad and up-to-date overview of the most relevant events at the local, national, and international levels.

It is important to note that Google News does not produce its own content and, therefore, is not a journalistic newsroom. It primarily uses automated algorithms to crawl and group articles under the same topic and from different sources with the aim of (1) offering multiple perspectives on the same news story, (2) detecting trends that may influence agenda setting, and (3) ranking them by relevance, timeliness, and source credibility (Support Google, 2025a).

The result is a kind of personalized digital newspaper front page. In fact, users can customize their experience in various ways, as they can specifically choose a wide range of topics, sources, locations, and interests (Support Google, 2025b). Additionally, Google adapts the content based on the user's search history and preferences, although it always prioritizes sources with a strong reputation which comply with its quality guidelines (Support Google, 2025a; 2025b).

Appearing on Google News can represent both an opportunity and a significant challenge for media outlets. On the one hand, it provides tremendous global visibility, which translates into a greater share of readers (Cordeiro, 2024); on the other hand, it forces media organizations to adapt to algorithmic logics that may favour large media conglomerates over smaller outlets—potentially endangering media diversity if not properly regulated or managed (Cordeiro, 2024).

As previously mentioned, appearing in Google News can serve as a mark of quality for a news outlet and can significantly boost its reach and site traffic. However, competition for

inclusion is intense and driven by algorithmic criteria (Support Google, 2025c). Under these circumstances, media organizations are compelled to optimize their content by following recommended best practices, which will be discussed throughout this chapter.

### 3. What is Google Discover: How it works and what it means for the media

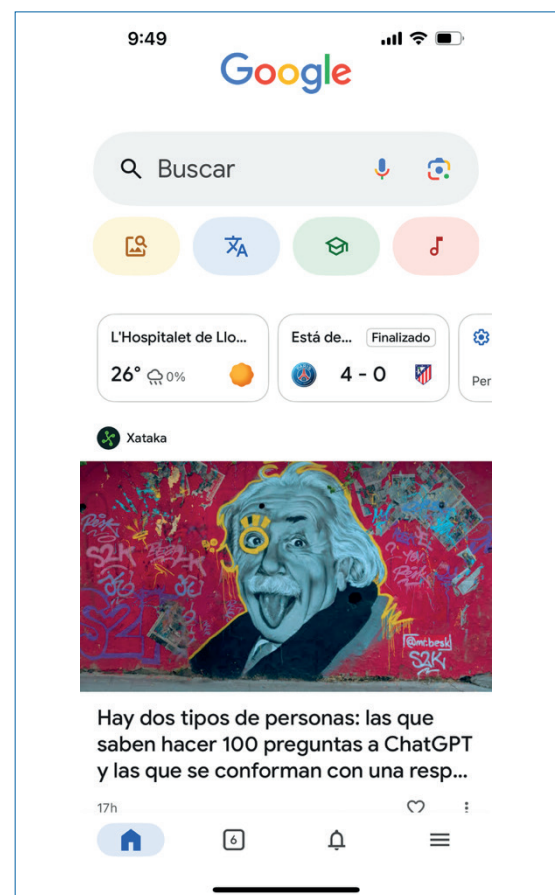
Google Discover is a service by Google that recommends content such as news (Absi-Flores & León-Ferreyros, 2024), articles, videos, or advertisements directly on the user's mobile phone, without the need for an active search (Lopezosa et al., 2022; Google Developers, 2020).

Unlike the traditional search engine, where the user types a query, Discover presents a personalized feed that appears automatically on Android devices, within the Google app and in the browser (Support Google, 2025d). Since mid-2025, the service has also begun to be available in desktop versions in some countries (González, 2025).

The content that appears in Google Discover is selected by algorithms, specifically using a machine learning technique called reinforcement learning, which optimizes recommendations in real time based on each user's behaviour (Strzelecki & Rizun, 2023): for example, what news the user has read before, how much time they spend on it, their activity across websites and apps, or their geographic location (Support Google, 2025d; Lopezosa et al., 2022). Thanks to this functionality, Discover is considered a tool for algorithmic curation (Lopezosa et al., 2022).

Although Google provides guidelines to prioritize useful, reliable, and user-focused content (Google Developers, 2025a), both industry analyses (Del-Castillo, 2024; Polo, 2023) and academic studies (Lopezosa et al., 2023) have noted a significant presence of sensationalist headlines and clickbait strategies in the feed, as aggressive adaptive tactics to optimize click-through rates (CTR). In response, Google has recently implemented updates and penalties aimed at reducing misleading and low-quality content (González, 2024), particularly following the core updates in March (Nelson, 2024) and August 2024 (Mueller, 2024).

**Image 2.**  
*Google Discover Front Page (06/16/2025).*



#### 3.1. How It works

Google Discover is a tool that automatically displays a selection of personalized content to each user, without requiring them to perform a keyword search. Instead of typing what they



want to find (as is done with the Google search engine), users encounter a feed of articles, news, and other formats that Google ranks based on thematic affinity, according to their recent browsing history, searches, or location (Support Google, 2025d) and algorithmic evaluation (Strzelecki & Rizun, 2023; Infobae, 2019).

User interaction with Discover is mostly passive, although at the time of writing, users can indicate whether they like a piece of content, want to see less of similar content, are not interested in a particular topic, do not want more content from a certain source, or even report and send feedback to Google about a specific result. This reinforces a one-way dynamic: it is Google that decides, based on its systems, what to show, and the user consumes content without actively participating in the process, as is the case on social media.

This feature forces journalists to adapt their content to the logic of engagement and contextual relevance rather than to a specific search engine positioning strategy (Lopezosa et al., 2022).

From the perspective of media outlets and professionals who manage their websites (editors, journalists, or webmasters), Discover represents a potentially valuable source of traffic. In many cases, it accounts for a significant portion of the visits received by a digital media outlet (Lopezosa et al., 2023).

However, achieving visibility on this channel is not easy: the system does not allow for registration or requests for inclusion, and the criteria that determine which content is displayed are not publicly known. Nevertheless, there are certain recommended practices that can help improve the chances of appearing: creating original content, using high-quality featured images, signing articles with real names, and ensuring what Google calls E-E-A-T: experience, expertise, authoritativeness, and trustworthiness (Google Developers, 2025b).

### 3.2. What It means for the media

Google Discover has direct implications for news media for several reasons. Firstly, it is a highly personalized source of web traffic, which can attract a large number of quality readers. Secondly, it can represent a growing share of organic traffic for media outlets that already have good positioning on Google.

Unlike search engine optimization, Discover means that media outlets do not compete for keywords, but rather for the user's interest through relevant and visually appealing content. Furthermore, strategies are prioritized that ensure content includes strong images, attractive and clear headlines, verified information, and the journalist's or author's expertise on the topic — rather than classic SEO strategies. That is, the ability to keep users engaged is more important than the newsworthiness of the topic.

Although there is no guaranteed way to appear in the feed, applying certain best practices — such as those discussed later — can improve visibility, especially those related to images, headlines, and the technical structure of the content (Lopezosa et al., 2023; Google Developers, 2025a).

These strategies to compete for attention in the feed (such as the use of striking headlines or images) can also affect content quality, although Google has implemented recent updates to reduce misleading and low-quality content. Other challenges also remain, such as traffic instability —since algorithm changes can cause a sudden drop in audience— and the difficulty of editorial planning, as it is hard to know exactly which factors the system prioritizes.

## 4. Recommendations for journalists and media outlets

Before offering a series of tips to journalists and editors on how to position news content in Google News and Google Discover, it is worth taking a moment to summarize the main features of both services to highlight their differences and similarities.

**Table 1**

*Features of Google News and Google Discover*

Characteristics	Google News	Google Discover
Main Focus	Current news and coverage of relevant events at the local, national, and global levels.	Personalized content based on user interests and behaviour, not limited to news.
Personalization	Personalization through selection of topics, sources, and locations; includes sections like "For You."	Automatic personalization based on search history, browsing activity, location, and interaction.
Type of Content	News from recognized media outlets, articles, videos, and full event coverage.	News, articles, blogs, videos, products, ads, or entertainment in a news-like format.
Feed Update	Periodic updates with the most relevant and recent news.	Dynamic, real-time feed that changes according to user behaviour and interests.
User Interaction	Follow/unfollow topics or sources, mark articles as "like" or "dislike."	Adjust preferences with options like "like," "show more/less," follow/unfollow topics (in some countries), and hide sources.
Coverage and Depth	Offers "Full Coverage" to display multiple perspectives on the same topic.	No "Full Coverage"; content is more varied and not grouped by topic.
Access and Availability	Mobile app (Android/iOS), web (news.google.com), integration with other Google products.	Google app, Chrome on mobile, google.com on mobile, home screens of some devices, and from 2025, on desktop in some countries.
Recommendation Algorithm	Based on selected interests and explicit user preferences.	Based on artificial intelligence and machine learning, analyzing multiple behavioural signals.
Purpose	Keep users informed about global and local current events.	Suggest relevant and engaging content without prior search, based on algorithmic prediction.

We now present a series of practical recommendations to improve visibility on Google News and Google Discover.

### 4.1. Ten best practices for optimizing the presence of digital media in Google News

1. Submit your news website to Google News via the Google Publisher Center. Before doing so, ensure that all technical and editorial requirements outlined in the guidelines are met.
2. Ensure transparency of the media outlet by publishing information about the journalists, editorial policies, and contact details.
3. Publish original content and avoid duplicate material.
4. Update the website frequently, maintaining a consistent publishing schedule for news and topic coverage.
5. Use consistent tags and categories to facilitate algorithmic classification by Google.



6. Comply with Google News content policies, especially regarding sensationalist headlines, misinformation, and offensive language.
7. Optimize the technical performance of the website, including fast load times and responsive design for mobile devices.
8. Include structured data using the schema.org standard to help Google interpret the content more effectively.
9. Optimize the robots.txt file to avoid crawl blocks that may prevent Google bots from indexing the site.
10. Adopt a clear editorial strategy that combines current news coverage with analytical depth, enabling content to appear in multiple Google News sections such as "International," "Local," "Economy," "Science & Technology," "Entertainment," etc.

#### 4.2. Ten best practices for optimizing news content in Google Discover

1. Publish news-style content, focused on either real-time trends or evergreen topics, that adds value and has viral potential.
2. Write engaging and accurate headlines that enhance visibility in a competitive environment without violating Google's policies.
3. Include large, high-quality, original images (at least 1200px wide) that are relevant to the content and properly credited.
4. Optimize author profiles and sign articles with real names to meet E-E-A-T criteria (Experience, Expertise, Authoritativeness, and Trustworthiness).
5. Apply SEO best practices, use structured data, ensure responsive design for mobile, and maintain fast page load speeds.
6. Publish frequently and enrich content with multiple formats including videos, tables, infographics, and lists.
7. Measure and analyze performance using Search Console. Use key metrics (impressions, clicks, CTR) to refine strategies and understand when and how Discover traffic is generated. Use tools like Google Trends or Discover Snoop to monitor trending topics outside the media outlet.
8. Drive traffic through social media, homepage features, hot topics, push notifications (WhatsApp, Telegram, etc.), and email newsletters.
9. Focus on thematic specialization and build authority, especially in YMYL (Your Money or Your Life) areas like health, finance, or science.
10. Comply with Google Discover's editorial policies: avoid penalized practices such as AI-generated content without human oversight, plagiarism, misleading headlines, lack of editorial transparency, or fake author profiles.

## 5. Conclusions

Nowadays, Discover and News are two essential pillars for improving the web visibility of online media. Both Google services have transformed both the way users access information and how digital media can distribute their content on a large scale.

Throughout this chapter, we have seen that Google News offers a structured and reliable channel for disseminating current news, allowing media outlets to position their brand and journalistic rigor before a global audience interested in accurate and relevant information. In turn, Google Discover has revolutionized news consumption: first, by incorporating an algorithmic logic that anticipates the reader's interest when deciding what to recommend, and second, by significantly boosting the potential audience of digital press and influencing newsrooms —where the tool's opaque operation and the difficulty of controlling visibility remain major challenges.

This work has sought to highlight the value of both platforms as tools to connect media outlets with broad and segmented audiences and, as a result, to drive readership growth and open new opportunities for monetization.

While the strategic use of Discover and News is crucial for media to consolidate their digital presence, we do not wish to conclude without acknowledging the challenge of applying these strategies with a critical eye toward the tension between algorithmic visibility demands and journalistic quality standards.

In this context, the challenge lies in applying these strategies to optimize the digital presence of media while keeping the reader at the centre —and without abandoning journalism's social commitments.

Therefore, it is up to journalists and editors to ethically adapt to constant algorithm changes, always safeguarding the integrity and core principles of journalism. This is the only way in which digital media will be able to rely on these two Google services without compromising their essence.

## 6. Funding

This work is part of the Project "Parameters and strategies to increase the relevance of media and digital communication in society: curation, visualisation and visibility (CUVICOM)". Grant PID2021-123579OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU.

## 7. References

Absi Flores, L. M., & León Ferreyros, A. (2024). Posicionamiento de contenido digital en Google Discover: el caso del diario peruano El Comercio. *Estudios sobre el Mensaje Periodístico*, 30(3), 625-634. <https://doi.org/10.5209/emp.96024>

Cordeiro, D. F., Lopezosa, C., Guallar, J., & Vázquez, M. (2024). Análisis de la cobertura de Google News: un estudio comparativo de Brasil, Colombia, México, Portugal y España. *Contratexto*, (42), 177-208. <https://doi.org/10.26439/contratexto2024.n42.7212>

Cordeiro, D. F. (2024). Perspectivas en contraste: análisis comparativo cuantitativo España y Brasil de la cobertura del conflicto israelí-palestino en Google News. *Documentación de las Ciencias de la información*, 47, 15. <https://revistas.ucm.es/index.php/DCIN/article/view/92187>

Del Castillo, C. (2024). El 'periodismo de algoritmo' ataca de nuevo: así es como titulares engañosos y supermercados inundan los medios. *elDiario.es*, 29 October. [https://www.eldiario.es/tecnologia/periodismo-algoritmo-ataca-nuevo-titulares-enganosos-supermercados-inundan-medios\\_1\\_11734185.html](https://www.eldiario.es/tecnologia/periodismo-algoritmo-ataca-nuevo-titulares-enganosos-supermercados-inundan-medios_1_11734185.html)

González, D. (2024). Google Discover tumba tu tráfico si titulas así. Te envía un mensaje como este. Es su nueva cruzada contra el clickbait. *Red de Periodistas*. <https://www.reddeperiodistas.com/google-discover-desplomara-tu-trafico-si-titulas-asi-y-te-envia-un-mensaje-como-este-en-su-nueva-cruzada-contra-el-clickbait>

González, D. (2025). Google Discover en escritorio llega en estos días, pero viene con truco. *Red de Periodistas*. <https://www.reddeperiodistas.com/google-discover-en-escritorio-llega-en-estos-dias-pero-viene-con-truco>

Google Developers. (2020). Aparecer en Discover. *Google Developers*. <https://developers.google.com/search/docs/appearance/google-discover?hl=es>

Google Developers. (2025a). Crear contenido útil, fiable y centrado en las personas. <https://developers.google.com/search/docs/fundamentals/creating-helpful-content>

Google Developers. (2025b). Experience, Expertise, Authoritativeness, and Trust (E-E-A-T). In *General Guidelines* (pp. 26-28). <https://guidelines.raterhub.com/searchqualityevaluationguidelines.pdf>

Guallar, J. (2015). Prensa digital en 2013-2014. *Anuario ThinkEPI*, 9, 153-160. <https://doi.org/10.3145/thinkepi.2015.37>

Infobae (2019). Cómo usar Google Discover para ver las noticias que te interesan en tu celular. *Infobae*. <https://www.infobae.com/america/tecnologia/2019/09/02/como-usar-google-discover-para-ver-las-noticias-que-te-interesan-en-tu-celular>

Iglesias-García, M., & Codina, L. (2016). Los cybermedios y la importancia estratégica del posicionamiento en buscadores (SEO). *Opción*, 32(9), 929-944.

Lopezosa, C., Codina, L., & Rovira, C. (2019). *Visibilidad web de portales de televisión y radio en España: ¿qué medios llevan a cabo un mejor posicionamiento en buscadores*. Barcelona: Universitat Pompeu Fabra. Departament de Comunicació, Serie DigiDoc-EPI, 4. <http://hdl.handle.net/10230/36234>

Lopezosa, C., Giomelakis, D., Pedrosa, L. & Codina, L. (2023). Google Discover: uses, applications and challenges in the digital journalism of Spain, Brazil and Greece". *Online Information Review*. <https://doi.org/10.1108/OIR10-2022-0574>

Lopezosa, C., Guallar, J., & Santos-Hermosa, G. (2022). Google Discover: entre la recuperación de información y la curación algorítmica. *SCIRE. Representación y organización del conocimiento*, 28(2), 13-22. <https://doi.org/10.54886/scire.v28i2.4796>

Lopezosa, C., Váñez, M., & Guallar, J. (2024). The vision of Google News from the academy: scoping review. *Doxa Comunicación*, 38. <https://doi.org/10.31921/doxacom.n38a1891>

Mueller, J. (2024). Qué debes saber sobre nuestra actualización principal de agosto del 2024 / Blog central de la Búsqueda de Google. *Google for Developers*. <https://developers.google.com/search/blog/2024/08/august-2024-core-update?hl=es> 419

Nelson, C. (2024, marzo 5). Qué deben saber los creadores web sobre la actualización principal de marzo del 2024 y las nuevas políticas de spam. <https://developers.google.com/search/blog/2024/03/core-update-spam-policies?hl=es>

Polo, S. (2023). Google Discover resucita la edad dorada del 'clickbait': "Prioriza lo más leído y eso es una ventana al sensacionalismo". *El Mundo*, May 16. <https://www.elmundo.es/television/medios/2023/05/16/6461f0fae4d4d8907a8b459c.html>

Salaverría, R., & García-Avilés, J. A. (2008). La convergencia tecnológica en los medios de comunicación: retos para el periodismo. *Trípodos*, (23), 31-47.

Strzelecki, S., & Rizun, M. (2023). Exploring the Impact of Google Discover on Users and Publishers: A Data-driven Study. *31° International Conference of information System Development*. <https://doi.org/10.62036/ISD.2023.28>

Support Google (2025a). *Empezar a utilizar Google News*. <https://support.google.com/googlenews/answer/9005669?hl=es&co=GENIE.Platform%3DAndroid>

Support Google (2025b). *Cómo empezar a usar Google*. <https://support.google.com/googlenews/answer/9005669?hl=es-419&co=GENIE.Platform%3DAndroid>

Support Google (2025c) *News content across Google*. <https://support.google.com/news/publisher-center/answer/9607025?hl=en>

Support Google (2025d). *Personalizar el contenido de Discover*. <https://support.google.com/websearch/answer/2819496?hl=es&co=GENIE.Platform%3DAndroid>

# Content curation in independent and media newsletters. Good practices and recommendations

**Pere Franch**

Universitat Ramon Llull, Spain

<https://orcid.org/0000-0002-4180-1168>

**Javier Guallar**

Universitat de Barcelona, Spain

<https://orcid.org/0000-0002-8601-3990>

---

Franch, P., & Guallar, J. (2025). Content curation in independent and media newsletters. Good practices and recommendations. In J. Guallar, M. Vázquez, & A. Ventura-Cisquella (Coords). *Digital communication. Trends and good practices* (pp. 25-36). Ediciones Profesionales de la Información. <https://doi.org/10.3145/cuvicecom.03.eng>

---

## Abstract

The recent surge in newsletters confirms their effectiveness as digital information products. Through content curation, newsletters enable journalists to connect with individual users specifically interested in their subject area. In this chapter, we present the main conclusions of six studies on newsletter curation from legacy and digital media outlets, as well as from independent journalists. We also provide a compendium of best practices and recommendations for improving the quality of journalistic newsletters. Offering quality newsletters compensates users for the overwhelming amount of information on the internet. At the same time, newsletters allow media outlets to strengthen their ties with their audience, increase user loyalty, and, eventually, increase their subscriber base. To improve quality, we recommend that newsletters have an identified author and abundant, curated content from varied sources in terms of origin (internal and external), time range, and morphology. There should also be variation in the use of curation techniques and link functions.

## Keywords

Content curation; Newsletters; Digital journalism; Media; Information sources; COVID-19; coronavirus; Independent editors.

---

## 1. Introduction

The newsletter is an old yet new communication channel that has experienced a significant boom in recent years. It is probably the most widespread and recognized means of curating digital content and has been the focus of several recent research studies.

This chapter summarizes the main conclusions of the CUVICOM research project (<https://www.ub.edu/cuvicom/>) on curated newsletters from media outlets and independent publishers. It also proposes recommendations and best practices for newsletters and content curation.

Since 2021, our research on curated newsletter content has primarily resulted in six articles, which focus on:

- Spanish legacy and digital native media newsletters (Guallar et al., 2021a).
- Spanish independent newsletters (Cascón-Katchadourian et al., 2022).
- Newsletters specializing in coverage of the SARS-CoV-2 pandemic (Guallar et al., 2022).
- Latin American independent newsletters (Lopezosa et al., 2023).
- The views of independent editors (Lopezosa et al., 2024).
- *The New York Times'* offer (Guallar et al., 2025).

These papers combine two methodologies: semi-structured interviews and analysis of newsletter samples using the CAS (Curation Analysis System) (Table 1). This method, initially described in Guallar et al. (2021b), observes two dimensions — content and curation — and studies the following parameters: for the first dimension, quantity, time range, origin (own or external) and sources (according to organization and morphology); for the second one, authorship, sense-making technique and link function.

## 2. Main research conclusions

### 2.1. Newsletters from Spanish newspapers

The first study, which focused on newsletters published by Spanish legacy and native digital media (Guallar et al., 2021a), concluded that, at the time the data was collected (in 2020), all media continued to prioritise links to their own content over external content. This consequently results in a limited variety of information sources and a surprisingly low presence of social media content, despite social media being considered a conducive channel for news curation.

In examining the characteristics of curation, the initial research established two major types of newsletter curation: intellectual and automated.

Intellectual curation is carried out by a professional whose authorship is identified. It uses one or more sense-making techniques and incorporates links for various informational purposes. In contrast, automated curation does not offer a journalistic contribution, nor is its authorship identified. It is based on a series of links to which no sense-making technique is applied, and there is no variety in the informational use of the links. This dual concept relates to Andringa's later contribution (2022), which distinguishes two types of newsletters: functional and relational. The functional model focuses on providing the audience with useful and relevant information. This model usually has a very specific audience and does not encourage a direct relationship between the author and subscribers. Most automated newsletters follow this model. The relational model, on the other hand, focuses on building relationships with the audience through personal content and engagement practices. These newsletters seek higher levels of interaction with their audience. For this model, it is indispensable that the authors leave their intellectual imprint.

The study also found that most newsletters publish a moderate amount of content, often around ten pieces per newsletter. The curated content is primarily current information published within the last few hours, almost exclusively from media outlets, mainly digital. Other sources, such as social media or official websites, have a limited presence.

### 2.2. Spanish independent newsletters

The study of Spanish independent newsletters (Cascón-Katchadourian et al., 2022) yielded the following conclusions: The main characteristics of these newsletters are that they are free, created by identified authors, and usually published weekly or more frequently. They cover a wide range of topics, from general current affairs to specialized information in specific areas and niches. They maintain a high level of audience loyalty, with open rates above 50% in many cases. Some have related products, such as podcasts.

In terms of the quality of their content curation, all of the studied newsletters use intellectual or professional curation, rather than automated curation. This is one of the main differences from the previous study.

It is also worth noting that there is diversity of temporal ranges in curated content, with a predominance of recent content (last few days), which is clearly related to the abundance of weekly periodicity among the analyzed newsletters, as well as a regular presence of retrospective and current content (last few hours).



Spanish independent newsletters also present a varied use of sense-making techniques, mostly combining two or three of them per bulletin. The “summarize” technique is present in all cases, and “quote” and “comment” are used frequently. Among the variety of link uses, the functions “describe”, “unchanged”, and “cite source of content”, in that order, stand out.

In terms of sources, independent newsletters mostly cite external content, revealing two emerging trends in their usage. First, some newsletters predominantly cite media websites, leaving little room for other sources. Second, others cite social media (social networks and blogs) more frequently.

This study included semi-structured interviews with the editors of the ten analyzed newsletters. Based on their responses, it was concluded that their primary concern is increasing the number of subscribers. They are reluctant to offer a paid product and state that the boom period of newsletters could lead to saturation and subsequent readjustment. The editors also believe that the future of content curation involves greater specialization and expanding products to different formats.

### 2.3. Coronavirus newsletters from newspapers in five countries

The paper on newsletters specialising in Covid-19 news from the mass media of France, Germany, Spain, the United Kingdom, and the USA, concluded that these newsletters share most of the characteristics observed in previous studies on general newsletters (Rojas-Torrijos & González-Alba, 2018; Guallar et al., 2021a). However, there was a slight improvement in all the observed indicators compared to general newsletters. There was a greater presence of authorship, different time ranges, a greater variety of sources, and a broader range of sense-making techniques.

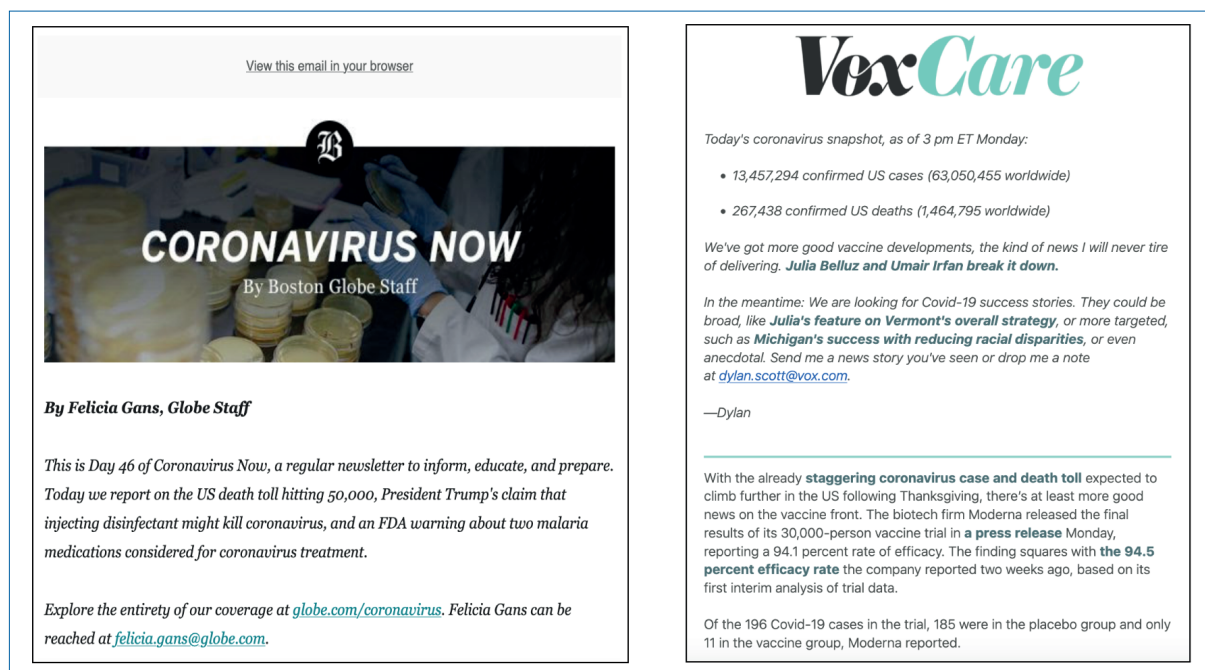
Legacy media outlets were found to offer little content from external sources. In contrast, pure digital media outlets offered the most external content, sometimes providing more links to external websites than to their own. It is also worth noting that all of the analyzed media outlets provided links to media websites and, to a much lesser extent, to official and corporate sources (often associated with health) and citizen sources. The origin of the content and format was most diverse in pure digital media such as elDiario.es, Politico, Vox, and BuzzFeed, which included not only websites, but also blogs and social media. When they offered content from other media outlets, it was usually from major international benchmark companies, including newspapers, television networks, and news agencies. Links to specialized content were, in most cases, to organizations such as the UN or WHO, or to expert medical websites. These results lead us to conclude that legacy media outlets that have transitioned to the digital environment primarily use this technology to increase webpage traffic. In contrast, pure digital media use hyperlinks and the web to provide users with access to a variety of opportunities, including social media.

Newsletters with automated content and a huge amount of links to their own website aim to increase traffic without personalizing the newsletter’s content. This trend has been observed more in legacy media than in pure digital media. While their purpose is informative, they also encourage subscriptions, website visits, micropayments, and content downloads to win and retain users (Silva-Rodríguez, 2021).



**Figure 1**

These are examples of newsletters about the SARS-CoV-2 pandemic from legacy and pure digital media outlets: The Boston Globe and Vox, respectively.



## 2.4. Latin American independent newsletters

This research provided an analysis of the main characteristics and quality of content curation in 12 independent newsletters in Latin America. It confirmed the effective use of content and curation dimensions by all the studied newsletters and identified interesting patterns. These include a significant amount of curated content per newsletter (between 11 and 47 elements), a greater use of timeless content together with recent or current content, a greater presence of content from external sources, and a predominance of sense-making techniques such as "comment" and "summarize".

All of the analyzed newsletters offer a high level of quality, as evidenced by their favorable position compared to other newsletters analyzed in previous studies. In all cases, the curation is elaborate, "intellectual," or authored, not automated.

The study reveals also three main groups of sources in the curation of Latin American independent newsletters: social networks (mainly Twitter, YouTube, and LinkedIn); media outlets from around the world or the Latin American region (e.g., *The Guardian*, *El País*, *O Globo*, *Folha*, *El Universal*, and *The New York Times*); and the newsletters themselves.

Figure 2

Examples of curation in two independent newsletters, from Latin America and Spain: *Lunes*, and *Africa Mundi*.



## 2.5. Interviews with editors

This paper, based on interviews with independent newsletter editors, reveals several key points:

Editors draw inspiration from online content, organizational tools, and curation methods, all of which influence the structure and design of newsletters. This aligns with other studies highlighting the importance of effective content curation and storytelling skills.

Editors perceive a positive rise in newsletters, attributing it to strategic aspects such as niche communities, original content, diverse monetization models, multimedia integration, and continuous improvement and expansion. Editors also emphasize the appeal of newsletters in providing personalized information and facilitating user-friendly production, as well as their ability to escape algorithmic control.

Newsletters are also highlighted for their ability to build brand image, specifically for their strategic importance in establishing direct connections, attracting subscribers, and complementing traditional media. While some interviewees anticipate that newsletters will remain niche products, their unique ability to offer close, even 'intimate' communication fosters loyalty to the editor's personal brand. Editors also recognize newsletters as effective tools for acquiring subscribers and as potential competitors to traditional media.

Finally, the interviewees acknowledge the crucial role of content curation in the face of information overload. They recognize the impact of technological evolution and support collaborative efforts among media entities to provide comprehensive content recommendations and establish a trustworthy digital identity.

## 2.6. The New York Times' newsletters

The study of *The New York Times'* newsletters concludes that most of them (73%) are open access. This reinforces the idea that the newsletters aim to expand NYT content, increase impact, and grow the audience and, consequently, paid subscriptions. High frequency and

regularity of publication are key to audience loyalty, as noted in relevant literature (Isaac, 2019; Kim et al., 2023; Mondéjar, 2024; Santos-Silva & Granado, 2019; Seely & Spillman, 2021).

Most newsletters feature authors' opinions or news, segmented by time, geography, language, or topic. Few follow a relational or hybrid model, as most are functional and limited to providing information without direct interaction with the user. While newsletter authors vary, subscriber-only newsletters typically have a more personal style, with journalists providing expert content directly and interactively.

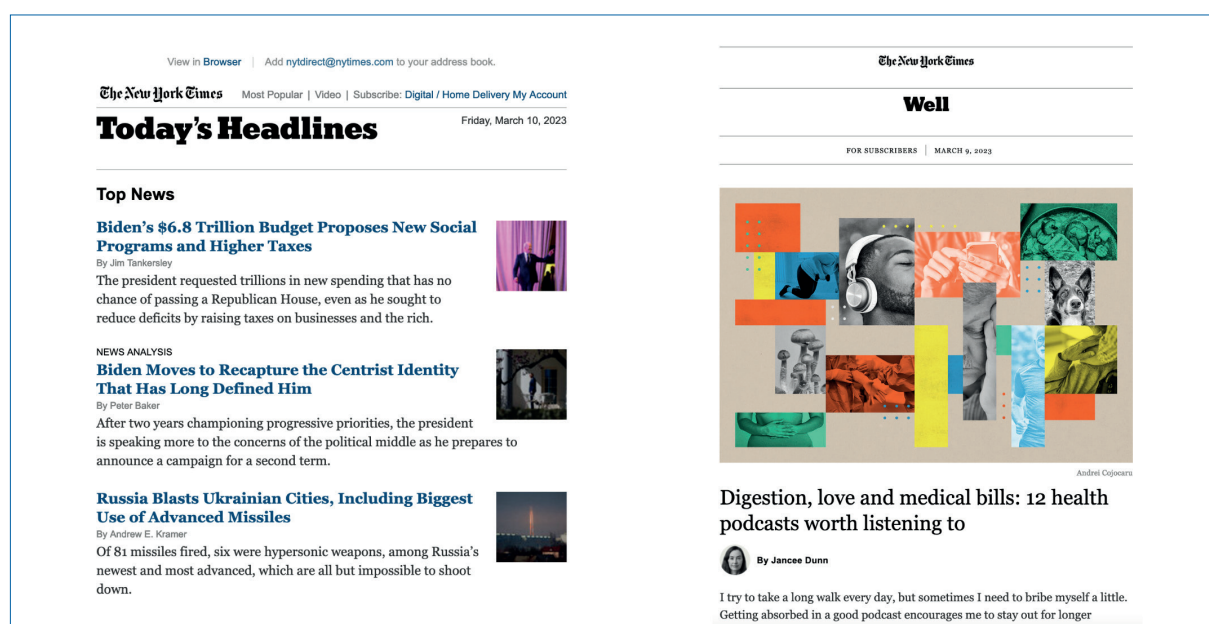
The amount of curated content is lower than in previous studies, especially in independent newsletters. There is more timeless and retrospective content, consistent with the informative and contextual nature of the NYT newsletters. Very little real-time content is present, reserved for following important and unexpected events. Most newsletters only offer content from *The New York Times*, which leads to increased traffic and better rankings. This raises the question of whether there is an endogamic tendency or corporatist impulse within mainstream media. In fact, media content is predominant in the links published. Newsletters with a specific author generally offer more external sources and are almost all subscriber-only. This external content is assumed to provide greater value to users.

In terms of the morphology of the curated content, websites predominate. There is also low variation in curation techniques, with "summarize" being the most frequent. "Unmodified" and "describe" are the two most used link functions, followed by "call to action".

The overall quality of NYT newsletters is lower than that found in other areas. Independent newsletters from Latin America received the highest scores (Lopezosa et al., 2023), and similar results were found in a study of Spanish newspaper newsletters (Guallar et al., 2021a). On the other hand, NYT newsletters with identified authorship received higher scores, corroborating previous studies on independent newsletters (Cascón-Katchadourian et al., 2022; Lopezosa et al., 2023). Finally, it can be assessed that the content curation of the NYT newsletters would be of average quality.

**Figure 3**

*Examples of NYT newsletters. On the left is an automated, non-authored, functional model newsletter. On the right is an authored, relational model newsletter with specific content, in this case, health and wellness.*



### 3. Best practices

The studies presented above confirm that newsletters are consolidating as a means of facilitating content curation, which can lead to an improvement in the quality of journalism. As a product personalized according to the user's specific interests, newsletters with quality curated content are an effective alternative to the saturated supply of material offered on the Web. Most of the studied newsletters contain specific content, thematically fixed, with hierarchical and time-limited information presented in a concrete format following the subscriber's preferences, who receives it in the privacy of his/her email (Carr, 2014; Pell, 2021; Rourke, 2021; Suárez, 2020).

Quality newsletters are a resource used by traditional media to strengthen ties with their audience, increase user loyalty, and, eventually, increase the number of subscribers (Isaac, 2019; Mondéjar, 2024; Santos-Silva & Granado, 2019). Their goal is to offer added value to pure and simple information — hard news — with which the author of the newsletter becomes a personal agent to the users, establishes a close (Andringa, 2022) or intimate (Mondéjar, 2024) relationship with them, and becomes a manager, coproducer and distributor of information, and at the same time adopts the role of networking creator (Upmanyu, 2024; Weder et al., 2023).

The following are some of the best practices for creating high-quality newsletters:

The CAS method (Table 1) evaluates newsletter quality based on the presence or absence of specific characteristics, so a more varied newsletter is considered higher quality. It is good practice to offer as much variety as possible across the parameters in all categories. This includes a generous amount of curated content, variety in time range, and balance between original and external content. Links to quality-curated content from diverse sources outside the medium are also valued. Source variety is valued when sources are not only media but also official, corporate, and citizen sources. Source type is valued when not only websites are offered but also blogs or social network content.

In terms of curation, an identified author is a sign of newsletter quality. The studies collected here establish a double dichotomy closely related to each other. First, a distinction is made between authored — intellectual — newsletters and automated newsletters. The value of the content and curation is generally considered higher in the former since it includes the added value of the author's intellectual work. Automated content curation, on the other hand, only provides links (usually from media outlets or the organization itself) to readers. Along these lines, Andringa (2022) introduces a similar taxonomy and distinguishes two types of newsletters. The relational type fosters a direct relationship between the author and the consumer, while the functional type can be the result of a good author's work and can provide valuable links, but without establishing this proximity or complicity.

A recommended practice for creating high-quality newsletters is to offer authored newsletters that follow the relational model, with content specifically aligned with the concrete interests of the target audience's niche.

The variety of sense-making techniques is valued among the other two curation categories, although the most commonly used technique is "summarize". Therefore, it is a good practice to present the newsletter content in a clear, concise summary so the reader knows what the product offers. And finally, variety in the functions is also a sign of quality in the text of the links. It should be noted that the most common function in most of the analyzed products

is “describe”. Consequently, a recommended practice is to present the text of the links in a manner that enables the reader to readily identify the content they will find on the linked website.

**Table 1**

*Dimensions, parameters and indicators observed by the CAS method*

Dimension	Parameter	Indicator
A. Content	A1. Number of contents	Quantity
		Retrospective or timeless information
	A2. Time range	Recent information
		Current information
		Real-time information
	A3. Origin	Own content
		External content
	A4. Source by type of organization	Official sources
		Corporate sources
		Media sources
		Citizens
		Websites
	A5. Source according to morphology	Blogs
		Social networks
		Secondary sources
B. Curation	B1. Authorship	Authorship
	B2. Sense-making technique	Summarize
		Comment
		Quote
		Storyboarding
		Unmodified
		Describe
		Contextualize
	B3. Link function	Interpret
		Cite source
		Cite author
		Call to action

Source: Prepared by the authors, based on Guallar et al. (2021b)

The use of internal or external sources merits further discussion. Previous studies have shown that legacy media generally offer more of their own content than external content and offer fewer links to external content than pure digital media.

In some cases, it was observed that legacy media outlets offered a high number of links to their own content. The Boston Globe Coronavirus newsletters are paradigmatic in this regard. This suggests that their objective may be to generate web traffic rather than to provide quality service to readers. In any case, this shows a self-referential attitude typical of the echo chamber effect denounced in the proliferation of social networks (Terren & Borge, 2021).

In contrast, digital native media outlets and independent newsletters offer more external content. Some even have no problem providing links to competing media, which we view positively because it enables users to broaden their knowledge and sources of information on topics that interest them.



## 4. Recommendations

Some practical implications in the form of recommendations for media, newsletter editors and content curation managers are presented in the following paragraphs:

### 4.1. Digital media outlets

- (a) Focus on quality content curation: Media outlets could improve their quality by offering newsletters with high-quality, curated content and more valuable external links.
- (b) Authored newsletters for connection: Newsletters with identified authorship and a personal style score higher in curation quality, so media outlets could promote these to connect better with the audience.
- (c) Thematic focus and segmentation: Publishing a range of newsletters with varied thematic segmentation can reach specific audiences; therefore, it is recommended that media outlets follow this strategy.
- (d) Web traffic strategy: The high presence of internal links, mainly in legacy media, suggests a strategy to boost web traffic, so media outlets could use newsletters to direct users to their websites and increase the number of paid subscribers.

### 4.2. Newsletters

- (a) The importance of curated content with added value: Successful newsletters offer curated content that provides context and analysis, in most cases using the “summarize” technique.
- (b) Include external links: While internal links are important for web traffic, external links from reliable sources add value and increase credibility.
- (c) Diversity of curation techniques: Most of the newsletters studied (except some of the independent ones) rely on a single technique, so using a variety of techniques, such as “comment”, “quote” or “storyboarding”, could enrich the user experience.
- (d) Hybrid model for greater connection: Combining the functional model of providing highly credited information with a relational approach that encourages audience interaction could improve newsletter quality and strengthen user relationships.

### 4.3. Regarding the use of content curation in journalism

- (a) Curation can be a tool to improve journalistic quality: Newsletters that effectively use content curation offer users a more comprehensive and enriching informative experience.
- (b) The journalist curator emerges as a manager and coproducer of information: The practice of content curation transforms the role of the journalist, who becomes a manager, co-producer, distributor of information, and also a network creator.
- (c) An alternative opportunity for freelance journalists: to create a personality as an expert in a niche topic, to publish alternative texts to traditional media, and to find an interim solution in the absence of other opportunities (Zilberstein, 2022).

## 5. Funding

This work is part of the Project “Parameters and strategies to increase the relevance of media and digital communication in society: curation, visualisation and visibility (CUVICOM)”. Grant PID2021-123579OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU.

## 6. References

- Andringa, P. (2022). *Email newsletters and the changing journalist-audience relationship*. PhD diss., The Oxford Internet Institute at the University of Oxford. <https://peterandringa.com/oii-thesis.pdf>
- Carr, D. (2014, June 29). For email newsletters, A death greatly exaggerated. *The New York Times*. <https://www.nytimes.com/2014/06/30/business/media/for-email-a-death-greatly-exaggerated.html>
- Cascón-Katchadourian, J.-D., Boté-Vericad, J.-J., & Guallar, J. (2022). Curación de información con marca de autor: análisis de newsletters independientes. *Ibersid: Revista de Sistemas de información y documentación*, 16(1), 41-53. <https://doi.org/10.54886/ibersid.v16i1.4799>
- Guallar, J., Anton, L., Pedraza-Jiménez, R., & Pérez-Montoro, M. (2021a). News curation by email: Analysis of the Spanish journalistic newsletters. *Revista Latina de Comunicación Social*, (79), 47-64. <https://doi.org/10.4185/RLCS-2020-1488>
- Guallar, J., Franch, P., Boté-Vericad, J.-J., & Anton, L. (2022). How do legacy and digital media curate coronavirus content. An assessment of newsletters from the USA and four European countries. *Profesional de la Información*, 31(3), e310318. <https://doi.org/10.3145/epi.2022.may.18>
- Guallar, J., Franch, P., Cascón-Katchadourian, J., & Boté-Vericad, J. J. (2025). The rise of curated newsletters in media: A case study of The New York Times. *Journalism Practice*, 1-22. <https://doi.org/10.1080/17512786.2025.2457332>
- Guallar, J., Pedraza-Jiménez, R., Pérez-Montoro, M., & Anton, L. (2021b). Curación de contenidos en periodismo. Indicadores y buenas prácticas. *Revista Española de Documentación Científica*, 44(2). <https://doi.org/10.3989/redc.2021.2.1742>
- Isaac, M. (2019, March 19). The new social network that isn't new at all. *The New York Times*. <https://nyti.ms/2AsoNcn>
- Kim, S. J., Kim, H., Choi, J. R., & Malthouse, E. C. (2023). Newly subscribed! Effects of e-mail newsletters on news-reading habit and subscriber retention during onboarding: Evidence from clickstream and subscription data. *Journal of Media Economics*, 35(3-4), 87-107. <https://doi.org/10.1080/08997764.2024.2333368>
- Lopezosa, C., Cascón-Katchadourian, J., & Guallar, J. (2023). Newsletters latinoamericanas de curación: estudio de casos y visión experta. *TransInformação*, 35. <https://periodicos.puc-campinas.edu.br/transinfo/article/view/10190>
- Lopezosa, C., Guallar, J., & Franch, P. (2024). Independent newsletters and content curation: The editors' view. *Tripodos*, (56), 167-182. <https://doi.org/10.51698/tripodos.2024.56.02>

- Mondéjar, D. (2024). "Newsletters. The renaissance of a valuable product to reach the audience." In *Innovations in Journalism*, edited by Klaus Meier, 241-248. London: Routledge. <http://doi.org/10.4324/9781032630410-30>
- Pell, D. (2021, November 10). The Internet's unkillable App. *The Atlantic*. <https://www.theatlantic.com/ideas/archive/2021/11/internet-newsletters-dave-pell/620664>
- Rojas-Torrijos, J.L. & González-Alba, J. A. (2018). La newsletter como producto periodístico en la búsqueda de nuevos lectores. Estudio de Boletines de Noticias de El País, El Español y El Independiente. *adComunica*, 15, 165–195. <https://doi.org/10.6035/2174-0992.2018.15.9>
- Rourke, M. (2021). The newsletter boom: capitalizing on the intimacy of the inbox. In: *Digital Content Next*. <https://digitalcontentnext.org/blog/2021/10/21/the-newsletter-boom-capitalizing-on-the-intimacy-of-the-inbox>
- Santos-Silva, D., & Granado, A. (2019). Old formats, new combinations: How newsletters represent innovation and blurring boundaries in digital journalism". In *ECREA Journalism Studies Conference 2019*, Vienna. <https://novaresearch.unl.pt/en/publications/old-formats-new-combinations-how-newsletters-represent-innovation>
- Seely, N., & Spillman, M. (2021). Email newsletters: An analysis of content from nine top news organizations. *Electronic News*, 15 (3-4), 123–138. <https://doi.org/10.1177/19312431211037681>
- Silva-Rodríguez, A. (2021). Emergencia de newsletters especializadas en Covid-19: Información curada y actualizada en el email. *Profesional de la Información*, 30(4), e300410. <https://doi.org/10.3145/epi.2021.jul.10>
- Suárez, E. (2020). How to build a successful subscription news business: lessons from Britain and Spain. *Journalist Fellowship paper*, Reuters Institute for the Study of Journalism. <https://bit.ly/3feaoPH>
- Terren, L. & Borge, R. (2021). Echo chambers on social media: A systematic review of the literature. *Review of Communication Research*, 9, 99-118. <https://doi.org/10.12840/issn.2255-4165.028>
- Upmanyu, K. (2024). Re-humanising news: Delineating the trends and research scope for curatorial journalism. *Studies in Media and Communication*, 12(1), 33–41. <https://doi.org/10.11114/smc.v12i1.6255>
- Weder, F., Weaver, C. K., & Rademacher, L. (2023). Curating conversations in times of transformation: Convergence in how public relations and journalism are "Doing" communication. *Public Relations Inquiry*, 12(2), 163-182. <https://doi.org/10.1177/2046147X231154550>
- Zilberstein, S. (2022). Digital platforms and journalistic careers: A case study of substack newsletters. *Columbia Journalism Review*. The Tow Center for Digital Journalism at Columbia's Graduate School of Journalism. [https://www.cjr.org/tow\\_center\\_reports/digital-platforms-and-journalistic-careers-a-case-study-of-substack-newsletters.php](https://www.cjr.org/tow_center_reports/digital-platforms-and-journalistic-careers-a-case-study-of-substack-newsletters.php)



# Content curation in social media for scientific journals. Trends and best practices

**Wileidys Artigas**

High Rate Consulting, USA

<https://orcid.org/0000-0001-6169-5297>

**Jesús Cascón-Katchadourian**

Universidad de Zaragoza, Spain

<https://orcid.org/0000-0002-3388-7862>

**Javier Guallar**

Universitat de Barcelona, Spain

<https://orcid.org/0000-0002-8601-3990>

---

Artigas, W., Cascón-Katchadourian, J., & Guallar, J. (2025). Content curation in social media for scientific journals. Trends and best practices. In J. Guallar, M. Váñez, & A. Ventura-Cisquella (Coords). *Digital communication. Trends and good practices* (pp. 37-48). Ediciones Profesionales de la Información. <https://doi.org/10.3145/cuvicom.04.eng>

---

## Abstract

This paper presents the main findings of several studies conducted within the Cuvicom project, which focuses on the social media presence of scientific journals in the fields of Communication and Library and Information Science. Special attention is given to their use of content curation and the social network X (formerly Twitter). The paper provides practical guidance in the form of trend analysis, recommendations, and best practices. It analyzes key aspects, such as the prevalence of content curation in social media posts, the topics of these posts, the specific curation techniques used, and the relationship between publications and types of publishers based on their size. The recommendations and best practices presented are expected to be useful for researchers in the field and for managers of scientific journals and social media.

## Keywords

Social media; Content curation; Scientific journals; X; Publishers; Communication; Information Science.

---

## 1. Introduction

In recent years, social media has become a strategic channel for scientific communication. Academic journals, aware of their potential, are using these platforms to varying degrees depending on the strategies of each publisher or editor, making them a subject of increasing research interest.

This chapter synthesizes the main best practices and recommendations derived from a series of studies conducted between 2022 and 2025 as part of the Cuvicom project. These studies focus on high-impact scientific journals in the fields of Communication (hereinafter, COM) and Information Science (LIS), both in *Iberoamerica* and globally. The data referenced is from the Scimago Journal Rank, SJR (<https://www.scimagojr.com/journalrank.php>), of the Scopus database. The paper addresses the adoption of social media, the use of curation techniques, the most frequent topics, and the strategies that generate the greatest impact. It also explores the relationship between social media use and the type of publisher based on size. The goal is to offer editors, journal managers, information professionals, and scholars a practical, useful, and applicable guide for optimizing their social media publications.

## 2. Social media adoption and use in scientific journals

### 2.1. Presence on social media of Q1 SJR COM and LIS journals

Data on the social media presence of scientific journals has been studied, and is presented in percentages because there are more COM journals than LIS journals, both in *Iberoamerica* (43 COM vs. 20 LIS) and among high-impact Q1 journals (112 vs. 61). As can be seen in Table 1,

there are significant differences between the two categories, depending on whether we focus on all Iberoamerican journals or global, high-impact journals, and always considering the data for active, proprietary profiles.

In Iberoamerican journals, LIS journals perform better than COM journals, whereas in global, high-impact journals, COM journals perform significantly better than LIS journals. In any case, while the evolution of journals in both categories tends to improve social media use, there is still ample room for improvement, particularly for Q1 journals in the LIS category.

The perception of this research is that Iberoamerican journals (which tend to have lower impact and appear in lower quartiles in the SJR) do not have much interest in uploading content to general social networks. In contrast, among global, high-impact journals, which are largely dominated by large and giant publishers, some publishers do not favor the use of proprietary social networks, but instead consolidate publications on the publisher's social media network.

**Table 1**  
*Journals and social media profiles.*

Category	Number of journals	Profile social network	Own and active profile
COM Latin America	43	20 (46.5%)	16 (37.2%)
LIS Latin America	20	12 (60%)	11 (55%)
COM Q1	112	100 (89.2%)	63 (56.2%)
LIS Q1	61	55 (90.16%)	16 (26.22%)

## 2.2. The role of small vs. large publishers

A study by Cascón-Katchadourian et al. (2024) analyzed the highest-impact journals in the SJR in both categories, examining the role of small, large, and giant publishers. The observed trends indicate that the vast majority belong to giant publishers (75% COM, 68% LIS), with Taylor & Francis being a notable leader. When combined with the fact that a significant number of the remaining journals belong to large publishers (19.6% LIS vs. 9.8% COM), we find that a total of around 85%-90% of the journals are from large publishers. These large publishers show poorer performance in the use and adoption of social media compared to small, independent publishers, which have a greater social media presence. The exception is the large publishers in the COM category, where 63.6% have at least one active social media profile. A possible hypothesis is that journals from small publishers tend to be more active on social media because they rely more heavily on visibility to stay competitive. In contrast, large publishers, with their high volume of publications and greater consolidation, may not require as much active presence on these platforms.

## 2.3. Preferred social network

Studies show a clear predominance of X (formerly Twitter) as the most-used social platform by scientific journals compared to other general social networks (Facebook, LinkedIn, Instagram, and YouTube). This trend has become more pronounced over the years. In 2022, Iberoamerican COM journals had the same number of active profiles on Twitter as on Facebook. However, by 2023, among Iberoamerican LIS journals, Twitter (now X) was used by 9 journals while Facebook was used by only 4. A 2024 study on the highest-impact LIS and COM journals

showed an even greater dominance of X over Facebook, both in the number of profiles and in exclusive profiles on the platform. Facebook appeared in a distant second place, but this network still holds a significant lead over Instagram, LinkedIn, and YouTube. Given these findings, the recommendation to prioritize X seems clear, although the notable presence of journals on Facebook should also be taken into account.

### 3. Curation levels

In general, all studies have shown that content curation in the social media posts of scientific journals has been high or very high. Among Iberoamerican communication journals, the level of curated posts is at least 80 percent for most of them (in 20 of the 24 profiles owned by the journals), with 9 profiles having 100 percent of their posts curated. A similar pattern is observed with Iberoamerican LIS journals, where 15 of the 18 profiles have more than 86% curated content.

Meanwhile, for global Q1 COM and LIS journals, we found that COM journals have 80% of their posts with curated content, while LIS journals surpass them with 93%.

It is also worth noting that, although studies on content curation generally focus on external curated content, scientific journals, like news media, tend to prioritize the curation of their own published content, which has also been called self-curation (Guallar et al., 2022).

### 4. The most used content curation techniques

In the study by Artigas and Guallar (2022) on Iberoamerican COM journals, it can be concluded that the most used technique is commenting, followed by summarizing. Citing and retitling are less common, while other techniques are virtually non-existent. Journals tend to predominantly use one technique in their curation style; for example, commenting is prevalent in *Review of Communication Research*, *Comunicación y Sociedad*, and *Revista Latina de Comunicación*, while summarizing is favoured by *Comunicar* and *Revista de Comunicación*, and citing is used by *Profesional de la Información*.

A year later, Cascón-Katchadourian et al. (2023) found that among Iberoamerican LIS journals, there were no instances of storyboarding or paralleling. With a few exceptions, each journal typically has a preferred technique it uses most often (for example, *EPI* prominently uses citing across its four social media networks). The analysis of social media profiles showed a certain balance between the use of commenting, abstracting, summarizing, and citing, with commenting being the preferred technique in six profiles, compared to the four profiles using the other techniques.

It is important to highlight that in terms of techniques, it is clear that each journal chooses and maintains techniques that it continues to use over time, whether they are individual or combined. This is evident from the repetition of the publishing format. Journals that use a combination of techniques have been able to leverage the added value of each technique and show different forms of curation. For example, they don't just copy the article title and authors to publicize the publication but also include citations or summarize content through videos or extracts to attract the reader.

Furthermore, a study of global Q1 journals in both areas (Cascón-Katchadourian et al., 2025) found similarities and differences. For instance, abstracting and summarizing were the main techniques used, while retitling was the least common. However, the study observed that the COM area uses summarizing more, whereas LIS makes greater use of abstracting, which highlights that the COM area uses more complex curation techniques.

The 2025 study also analyzed the method of content integration, observing the use of mentions (an average of one per post in both categories) and hashtags. LIS journals used an average of 2.68 hashtags per post, more than double the average for COM journals (1.46 per post), indicating that LIS journals make greater use of hashtags as a tool for categorizing posts on social media.

It can therefore be concluded that for Iberoamerica, the commenting technique is used most often (both LIS and COM), whereas when we look at the study of Q1 journals in the same areas, we see greater use of abstracting and summarizing. Table 2 provides a summary of the main findings.

**Table 2**  
*Techniques and their use in different publications*

	2022 (Iberoamericanas COM)	2023 (Iberoamericanas LIS)	2025 (Q1 LIS y COM)
Most commonly used technique	Comment	Comment	Abstracting (LIS) and summarize (COM)
Least commonly used technique	Rename	Rename	Retitle
Frequent combination	Summarize and comment	Abstracting and comment	Abstracting/summarize with commentary
New inclusions in studies	Not applicable	Not applicable	Integration of content, mentions, and hashtags

Source: Artigas & Guallar (2022), Cascón-Katchadourian et al. (2023; 2025).

## 5. Frequent topics

The data on the topics of social media posts by scientific journals is quite consistent across the published studies, with a few nuances. This data was collected using different methodologies, as the first two studies distinguished only three topic categories, while the last study in 2025 subdivided the “other” category.

In any case, curation prioritizes the journal’s own content, especially from the latest issue. For this reason, the most common topics revolve around the dissemination and promotion of recent articles. In most cases, it is a post presenting an article published by the journal. Additionally, we find other types of content, such as podcasts with authors, videos related to articles, and open invitations to review certain articles. In the article by Artigas and Guallar (2022), there is a tie between the “latest issue” and the “announcements and others” topics, due to the broad nature of the second category. In the study by Cascón Katchadourian et al. (2023), where the “announcements and others” category was not yet divided, the “latest issue” category was still the majority in 12 of the 18 profiles studied. Quantitatively, it was by far the most numerous category, with 275 of the 454 analyzed items of content (60% of the total). In the third study from 2025, for both the COM and LIS categories, “latest issue” was the main category, with 64.97% and 71.63% of the analyzed posts, respectively.

In second place, in the 2022 and 2023 studies, the “other/announcements or other topics” category encompasses various topics such as: Calls for papers, announcements of a new issue, themes of future issues, dissemination of the journal’s own activities, dissemination of external content, prize calls, journal rankings, and others. This is the second most important thematic category, ahead of “previous issues,” which consists of the dissemination of publications from issues prior to the current one. In the first study on Iberoamerican communication journals, this was the dominant category in 11 profiles from 8 journals, similar to what was seen with the latest issue. In the 2023 study, it was the predominant category in 5 profiles from 5 different journals and, quantitatively, the second most-used category with 25.5% of the curated content.

In the 2025 study, this category was divided into several subcategories with, logically, more modest percentages of posts. These subcategories included calls for papers, current volume dissemination, external content, journal’s own activities, and others.

**Table 3**  
*Other topics, excluding latest published issue.*

Curation themes	COM % of total posts	LIS % of total posts
Call for Papers	4.87%	2.02%
Current volume distribution	8.60%	2.77%
External content	2.33%	2.35%
Own activities	4.26%	5.40%
Others	3.44%	0.09%
Previous volumes	11.59%	15.71%

Source: Cascón-Katchadourian et al. (2025).

Therefore, the dissemination of articles from previous volumes, which was the smallest category in earlier studies, now emerges in the 2025 study as the second most important thematic category, with 11.59% of posts in COM and 15.71% in LIS, just behind the current issue.

It should be noted that this thematic category has grown in importance over the years. In 2022, it was the predominant category in only two profiles (one of which had a single post). In 2023, it was dominant in only one profile — the Twitter account of the *Revista Española de Documentación Científica*. Quantitatively, it also grew slightly from 13.8% of posts in LIS in 2023 for Iberoamerican journals to 15.71% in LIS in 2025 for high-impact journals.

## 6. Best practices in content curation techniques and topics

Finally, some cases of efficient and notable curation practices have been extracted and presented from the articles by Artigas and Guallar (2022) and Cascón Katchadourian et al. (2023). For each example, the social media post is shown with its essential characteristics (text, image description, link to the curated content), followed by a brief analysis.

## 6.1. Techniques

### Abstracting

#### *Revista Latina de Comunicación Social*

Text: 'La Noticias falsas y desinformación sobre el Covid-19: análisis comparativo de seis países iberoamericanos' escrito por Liliana Gutiérrez, Patricia Coba y Javier Andrés Gómez en el número 78 #Especial #Covid19 de #RevistaLatina

Image: Self-created graphic of a world map with the Iberoamerican area highlighted and the words Covid-19.

<https://twitter.com/revistalatinacs/status/1337412573839446019>

This is an example of abstracting, a technique characterized by its simplicity and often by automation. Although it offers little added value, it can allow a journal to maintain a minimal content presence on social media with little effort.

### Retitling

#### *Revista de Comunicación*

Text: "¿Netflix es entretenimiento, o hay otros elementos que van más allá del disfrute personal de los jóvenes? Lee el artículo de Benavides Almarza, C., & García-Béjar, L. <https://doi.org/10.26441/RC20.1-2021-A2> #rcom #udep".

Image: A photograph of a family watching Netflix on a television.

<https://twitter.com/RevComUdep/status/1370393863878279172>

In this case, the original title, which begins with a question ("Why do people who watch Netflix watch Netflix?"), is replaced in the tweet with another question ("Is Netflix just entertainment, or are there other elements that go beyond young people's personal enjoyment?"). This is an interesting use of retitling to pique the audience's interest in the article's content.

### Summarizing

#### *Revista Española de Documentación Científica*

Text: "Sergi Cortiñas y Marc Darriba Zaragoza estudian el porcentaje de obras pseudocientíficas presentes en #bibliotecas públicas. Resultados: mayor porcentaje #astrología | Más números de títulos sobre #homeopatía en proporción a #quimioterapia. <https://doi.org/10.3989/redc.2018.1.1474>".

Image: none.

[https://twitter.com/Revista\\_REDOC/status/1529795736191344640](https://twitter.com/Revista_REDOC/status/1529795736191344640)

This tweet is a great example of the summarizing technique. It's concise and direct, with all the content in the text itself and no image. It briefly summarizes the article's topic and two of its most significant findings, also mentioning the authors' names. Additionally, it uses hashtags precisely to identify keywords: libraries, astrology, homeopathy, and chemotherapy.



## Commenting

### Comunicar

Text: "We invite you to read this interesting paper, which describes a language learning model that applies social media to foster contextualized and connected language learning in communities. Read more on: <https://doi.org/10.3916/C50-2017-01#AcademicTwitter> #phdchat".

Image: Text with the article's title, authors, and their country, alongside a photograph showing a woman and two children looking at an illustrated book.

[https://twitter.com/Rev\\_Comunicar/status/1445812347998773253](https://twitter.com/Rev_Comunicar/status/1445812347998773253)

In this example, the tweet goes beyond just describing the article's content, using a more personal approach to the reader ("We invite you") and an opinionated, recommending tone ("this interesting paper").

## Citing

### Cuadernos.info

Text: "#LecturaRecomendada @NekaneParejo analizó 11 películas de Almodóvar centrándose en la figura del fotógrafo. Concluyó una evolución hacia la caracterización de otros roles: voyerismo, fotografía doméstica e imágenes robadas. Lee el artículo completo en <https://doi.org/10.7764/cdi.52.36267>".

Image: A highlighted quote from the article in quotation marks, with the specific page and author indicated. It also includes the journal's name, title, year of publication, issue number, and page numbers.

<https://twitter.com/CuadernosInfo/status/1564650714395287554>

This tweet is a good example of the citing technique. The quote is incorporated into the post not in the text, but through a well-designed graphic image. The selected quote highlights one of the most notable findings of the research, condensing the distributed article into a few words.

## 6.1. Topics

### Current issue

#### Profesional de la información

Text: "Thanks for joining our life: Intimacy as performativity on YouTube parenting vlogs. By @aVizcainoVerdu, @Patriciadcm and @DaniJaraDent #SocialMedia #Intimacy #Performativity #YouTube #Platforms #Vlogs".

Image: Includes the article title, names and portraits of the three authors, a direct quote from the article, and its reference data.

[https://twitter.com/revista\\_EPI/status/1556520670766456832](https://twitter.com/revista_EPI/status/1556520670766456832)

[https://twitter.com/revista\\_EPI/status/1559208973722583044](https://twitter.com/revista_EPI/status/1559208973722583044)

On Twitter (but not other platforms), EPI publishes two tweets for each new journal article: the first one appears as soon as the article is available in the continuous publication model, and the second one appears a few days later. The content of both tweets is mostly identical, but the second one is translated into English and includes a selected quote, as well as a different graphic format in the image. This duplication is considered a content recycling practice in the social media field, which is recommended, especially on platforms where content has a shorter useful life, such as Twitter (Sanagustín, 2017). This increases the chances of visibility and access to the new journal article.

## Previous issues

### Review of Communication Research

Text: “‘Social Norms: A Review’, authored by Adrienne Chung & Rajiv Rimal has been cited 100 times in Google Scholar. Congrats to the authors! Congrats also to the reviewers, & Lijiang Shen, who served as editor, for helping the authors to publish an excellent literature review. We are happy to see that the articles we publish are useful to scholars worldwide! <https://buff.ly/2CH5Kbc>”

Image: Group of people celebrating.

<https://www.facebook.com/RCommunicationR/posts/2792781654090639>

This is an excellent example of a post referencing an article published in a previous issue of the journal, informing the audience that it has reached a considerable number of citations on Google Scholar. The text includes references to the content and its authors and editor. (Note: The internal link from the Facebook post to the article was later broken after its selection for this study).

## Other topics

### BiD

Text: “Anàlisi de l’ISBN a Iberoamèrica <https://ub.edu/blokdebid/ca/content/analisi-de-lisbn-iberoamerica> amb Isabel-Cristina Arenas al #BlokdeBiD @revistaBiD”.

Image: Shows the image presented on the blog.

<https://twitter.com/revistaBiD/status/1506627616173965320>

A good example of a social media posts not strictly related to the journal’s own content are the regular tweets on BiD’s profile about the “Blok de BiD,” a related blog that publishes reviews of reports and studies on the journal’s topics. This practice can increase the visibility of its Twitter profile and, by extension, the journal’s brand.

## 7. Practical recommendations for scientific journals

Here are some recommendations for editors and managers of academic journals.

On adoption and use:

- Create dedicated social media profiles: Editors should consider creating specific social media profiles for each journal as a strategy to increase their visibility and impact.

- Maintain an active presence on a key platform: It's recommended to maintain an active presence on at least one main social network. X (formerly Twitter) is currently the preferred platform for the scientific community. However, given recent changes and how platforms may evolve, editors could also explore Bluesky. It's also a good idea to have a complementary social network, such as Facebook or LinkedIn. Ultimately, the best approach is to choose at least one social network where your target audience has the strongest presence. Editors should also consider using other platforms like Instagram, LinkedIn, or YouTube if they have a clear strategy and the resources to maintain them.
- Publish regularly: To avoid long periods of inactivity, journals should publish consistently. Having a social media profile to post only twice a year is not effective and creates a negative impression.
- Designate a responsible team: Assign a person or an editorial team to manage social media. Investing in content dissemination is a strategy that consistently yields positive results.
- Adopt diverse dissemination formats: Another recommendation is to use different formats to ensure a continuous presence. For example, journals like *Mediterránea* have successfully developed a media ecosystem that includes social media, audio summaries (podcasts), webinars, and a WhatsApp group (Guallar, 2024).

#### On Topics:

- Vary your content: It's recommended to publish content on a variety of topics. While new articles should be a priority, you can also give old publications a new life by curating all content — including external material — that might interest your community. Repurposing past content with fresh angles is a cost-effective and efficient way to add value to your journal's content.
- Foster community: Use your profile to build a community, not just to promote your own publications. Encourage users to visit your profile regularly for news in their specific field.
- Explore new themes: One innovative approach has been seen in journals that form alliances or host contests for the best papers, which allows them to interact with other journals at their level and attract a larger audience.
- Include external content: It is also recommended to include relevant and interesting external content for your target audience, such as news and current events related to your field.

#### On Techniques:

- Combine curation techniques: Use a mix of different curation techniques to spark reader interest.
- Integrate content effectively: Integrate curated content in various ways and use hashtags and mentions strategically to increase visibility.

## 8. Conclusions

Content curation on social media is a growing practice among scientific journals, according to studies conducted in recent years. Its ability to connect with audiences, synthesize the value of published knowledge, and expand the impact of articles positions it as a necessary practice in the current academic digital ecosystem.

The evidence gathered from recent research shows that, while there is still ample room for improvement, many journals are making use of curation on social media. From choosing the right platforms to applying editorial techniques that add value, content curation has proven to be an adaptable, scalable, and effective practice.

The recommendations and case studies analyzed in this chapter offer a concrete guide for other journals to replicate and adapt these practices to their specific needs. The key lies in understanding that the goal is not just to disseminate information but to engage in dialogue, interpret, and connect. This requires adopting an active, creative, and reader-centered editorial perspective.

Based on these findings, it can be affirmed that a coherent and sustained content curation strategy on social media is crucial for scientific journals. This practice not only provides visibility and dissemination for specific publications but also strengthens the journal's brand identity with a wider audience. It also offers authors a valuable platform to promote their work, increasing their chances of being read, evaluated, and eventually cited. In this sense, content curation by academic journals on their social channels fosters greater reader attraction through diverse presentation formats, which translates into superior impact and visibility for both the publications and their authors.

Furthermore, it would be important to explore, from the different perspectives of journals, which formats are most likely to generate interaction with the target audience based on their own contexts, realities, and interests. This remains a task for future research.

## 9. Funding

This work is part of the Project "Parameters and strategies to increase the relevance of media and digital communication in society: curation, visualisation and visibility (CUVICOM)". Grant PID2021-123579OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU.

## 10. References

Artigas, W., & Guallar, J. (2022). Curación de contenidos científicos en medios sociales de revistas Iberoamericanas de Comunicación. *Revista de Comunicación*, 21(2), 15–32. <https://doi.org/10.26441/RC21.2-2022-A1>

Cascón-Katchadourian, J., Artigas, W., & Guallar, J. (2023). Curación de contenidos en las redes sociales de revistas de información y documentación de Iberoamérica. *BiD: textos universitarios de biblioteconomía i documentació*, 51 (diciembre). <https://doi.org/10.1344/BID2023.51.05>

Cascón-Katchadourian, J., Artigas, W., Arroyo-Machado, W., & Guallar, J. (2024). Use of social networks by high-impact journals in Communication and Information Sciences. *Revista Mediterránea de Comunicación*, 15(1), 19–38. <https://doi.org/10.14198/MEDCOM.25487>

Cascón-Katchadourian, J., Guallar, J., & Artigas, W. (2025). Social media analysis of high-impact Information and Communication journals: Adoption, use, and content curation. *Publications*, 13(1). <https://doi.org/10.3390/publications13010005>

Guallar, J. (2024). 25 recomendaciones para revistas científicas en redes sociales. In *13° Congreso Internacional sobre Revistas Científicas CRECS*, Arequipa, Perú, 10 mayo 2024. <http://eprints.rclis.org/45758>

Guallar, J., Franch, P., Boté-Vericad, J.-J., & Anton, L. (2022). How do legacy and digital media curate coronavirus content. An assessment of newsletters from the USA and four European countries. *Profesional de la información*, 31(3), e310318. <https://doi.org/10.3145/epi.2022.may.18>

Sanagustín, E. (2017). *Plan de contenidos en medios sociales*. Editorial UOC.

# Curation of sources and equality narratives in the media agenda

**Anna Ventura-Cisquella**

Universitat de Barcelona, Spain

<https://orcid.org/0000-0002-3677-8055>

**Javier Guallar**

Universitat de Barcelona, Spain

<https://orcid.org/0000-0002-8601-3990>

---

Ventura-Cisquella, A., & Guallar, J. (2025). Curation of sources and equality narratives in the media agenda. In J. Guallar, M. Vázquez, & A. Ventura-Cisquella (Coords). *Digital communication. Trends and good practices* (pp. 49-68). Ediciones Profesionales de la Información. <https://doi.org/10.3145/cucom.05.eng>

---

## Abstract

This is an análisis of the work of equality journalists in the Spanish digital press, with an emphasis on the curation of documentary sources and content topics. The study is based on a corpus of 316 articles published during the first half of 2024 in two traditional newspapers (*El País* and *El Periódico*) and four digital native outlets (*eldiario.es*, *El Salto*, *Infolibre*, and *Público*), all of which currently employ journalists specialized in equality. The identification of topics was carried out through exploratory reading of headlines and full texts, followed by a frequency analysis of tags and keywords in each article, organized into a codebook. For source analysis, the CAS method (Curation Analysis System) was applied, which allowed classification of all curated links in each article according to type of organization: media outlets, official sources, corporate sources, or citizen sources. The results show that journalists perform a rich curation of documentary sources to support reports, craft headlines, and highlight inequalities. Four major themes were identified: gender inequalities and violence; advances and tensions in feminist movements and reproductive rights; monitoring of the Ministry of Equality and its policies; and progress and challenges regarding rights and violence toward the LGBTQ+ community. This study highlights the strategic role of these professionals in consolidating rights based narratives within the media agenda and documenting equality policies in Spain.

## Keywords

Content curation; Documentation; Gender editors; Digital press; Equality; Media agenda; Documentary sources.

---

## 1. Introduction

In recent decades, there has been a decline in the prominence of traditional media in favour of the major technological giants. This phenomenon has contributed to the emergence of an information ecosystem marked by the rise of “platform capitalism,” the proliferation of pseudo-media outlets, and the spread of hate speech that denies inequalities on social networks (Pan et al., 2025; García-García, 2025).

In light of this reality, it is essential to analyze how media content aimed at promoting gender equality is produced and disseminated, with equality understood as fairness in rights, responsibilities, and opportunities between women and men, as well as among girls, boys, and people with diverse gender identities (UN Women, 2025).

In this context, the work of journalists specializing in equality currently faces multiple challenges. These professionals are frequently questioned for curating and editing news content with a rights-based and gender-equality perspective — particularly when working for progressive media outlets — and mainly by ultra-conservative sectors (Bennett & Livingston, 2018; Reuning et al., 2022).

There is also a growing distrust toward official sources in Spain. This skepticism is instigated and amplified by certain far-right political groups, such as Vox. The situation complicates the practice of journalism specialized in equality, within a context governed by a progressive executive.



The erosion of trust in institutions, in turn, weakens the credibility of official information and of traditional media outlets where equality journalists work. This fuels the search for unverified “alternative” sources.

These alternative sources are often associated with far-right strategies such as the “manosphere,” a network of online communities focused on discussions of masculinity, frequently critical of feminism and traditional media (García-Mingo & Díaz-Fernández, 2023).

Finally, the authority of these professionals is also undermined by their being women journalists, as noted by various international organizations (UNESCO, 2019; Amnesty International, 2024).

In this scenario, documentary curation journalism has become a key strategy to ensure quality and rigour in reporting (Guerrini, 2013; Guallar, 2014; Cui & Liu, 2017; Lopezosa et al., 2023). This journalistic work is conceived as a set of practices aimed at selecting, linking, and contextualizing verified sources, with the goal of providing accurate, accessible, and socially relevant information (Guallar & Codina, 2018).

The present study analyzes the source-linking practices of equality journalists in the Spanish digital press. This object of study was chosen because, although the rise of social networks and other digital platforms has partly displaced the centrality of traditional media (Waisbord, 2018; Bailo, 2021), digital press still plays a key role as an agent in shaping the public agenda (Virgili & Puche, 2019). This makes it a crucial space for influencing social perceptions as well as the formulation of policies and legislative frameworks related to gender equality.

Moreover, the content of the digital press is not limited to its own environments, but also circulates widely across social networks. Even among those who do not turn to these platforms for informational purposes, incidental exposure to news is common and is associated with increased online news consumption (Feezell, 2017; Fletcher & Nielsen, 2017).

On the other hand, although previous research has addressed the role of equality journalists in newsrooms — focusing on aspects such as the inclusion of gender editors, their functions, and their ability to influence editorial lines across different sections of the media outlet (Spinetta, 2022; Iranzo-Cabrera et al., 2023) — there is still a lack of in-depth analysis of the content produced by these professionals. By examining the linking of documentary sources, this study explores how they seek to position headlines and materials related to gender inequalities and violence, as well as women’s and LGBTQ+ rights, within the Spanish media agenda. This work may be of interest to journalists and researchers seeking strategies to promote equality focused content and counter hate speech in the media.

## 2. Materials and methods

The content analysis is based on a corpus of 316 articles published by gender journalists during the first half of 2024, in two traditional Spanish newspapers (*El País* and *El Periódico*) and four digital-native outlets (*ElDiario.es*, *El Salto*, *Infolibre*, and *Público*), which are currently the ones that have these positions.

Of the six media outlets analyzed, three — *ElDiario.es*, *El País*, and *El Salto* — have gender editors who hold strategic decision-making positions within the management team. These professionals supervise, manage, and improve gender representation in the content across all sections of the outlet. The other three media outlets — *El Periódico*, *Público*, and *Infolibre* —

have heads of the equality section but they do not currently act as gender editors in the strict sense. Their work is mainly focused on creating and writing specific content for the equality section.

A sample of articles by gender journalists published in the first half of 2024 was selected from these outlets, a period marked by significant events. During this time, Spain ranked as the fourth-best performing European country in terms of gender equality (EIGE, 2023), and the Ministry of Equality transitioned from Irene Montero to Ana Redondo (Ministry of Equality, 2023), both within the progressive political spectrum but with different approaches to managing the portfolio. In this period, debates on trans rights, LGTBQ+ issues, and gender equality intensified, alongside the rise of denialist discourses from politics, youth, and social media (CIS, 2024; Europa Press, 2024).

Table 1 presents the six media outlets analyzed and the journalists responsible for the equality sections. It includes their roles, links to their author profiles, and the total number of articles published between January and June 2024. These articles make up the study sample.

**Table 1**

*Media outlets with gender journalists and their published content*

Newspaper	Name	Role	Author profile	Articles
El Diario.es	Anna Requena	Gender editor	<a href="https://www.eldiario.es/autores/ana_requena_aguilar/">https://www.eldiario.es/autores/ana_requena_aguilar/</a>	53
El País	Isabel Valdés	Gender editor	<a href="https://elpais.com/autor/isabel-valdes-aragones/">https://elpais.com/autor/isabel-valdes-aragones/</a>	48
El Periódico	Patricia Martín	Head of section	<a href="https://www.elperiodico.com/es/autor/patricia-martin">https://www.elperiodico.com/es/autor/patricia-martin</a>	29
El Salto	Patricia Reguero	Gender editor	<a href="https://www.elsaltodiario.com/autor/patricia-reguero">https://www.elsaltodiario.com/autor/patricia-reguero</a>	31
Infolibre	Sabela Rodríguez	Head of section	<a href="https://www.infolibre.es/autores/sabela-rodriguez/">https://www.infolibre.es/autores/sabela-rodriguez/</a>	76
Público	Candela Barro	Head of section	<a href="https://www.publico.es/author/candela-barro">https://www.publico.es/author/candela-barro</a>	79
Total				316

The methodology was an exploratory reading of headlines and full texts to analyze the themes. This was followed by a frequency analysis of tags and keywords used by gender journalists to classify and position each article. The results were organized into a codebook following Braun and Clarke (2006) and Krippendorff (2018). The CAS (Curation Analysis System) method was applied to analyse documentary sources, as it had already been used in studies on journalistic curation (Guallar et al., 2021; Lopezosa et al., 2023; Guallar et al., 2025).

Outgoing links from each article were classified according to the type of organization: media outlets, official sources, corporate sources, and citizen sources.

During the analyzed period, four main themes were identified: (1) gender inequalities, discrimination, and violence; (2) advances, setbacks, and tensions in feminist movements, with a focus on reproductive rights; (3) monitoring of the Ministry of Equality and the implementation of its policies and laws; and (4) progress and challenges regarding the rights and violence faced by the LGTBQ+ community.

In addressing these themes, documentary sources — media outlets, official documents, corporate reports, and tweets or citizen statements — were central. They were used both to craft headlines and support reports, as well as to analyze the construction of public debates and the monitoring of equality policies. The analysis was structured into five sections as follows:

- Some Characteristics of Source Curation
- How to Curate Gender Inequalities and Violence
- How to Curate Sources and Resources from Feminist Movements
- How to Curate Equality Laws and Policies
- How to Curate Advances and Setbacks in LGTBQI+ Rights

### 3. Some characteristics of source curation

During the first half of 2024, gender journalists linked to over two thousand documents (external links), focusing mainly on outlets such as El País and Público, followed by Infolibre, ElDiario.es, El Periódico, and El Salto. This demonstrates that gender-perspective curation is concentrated on a few media outlets that aim to influence the news agenda.

Regarding the type of linked sources, El País and Infolibre primarily rely on official sources to cover violence, inequalities, and public policies, while El Salto prioritizes community perspectives and feminist movements, curating corporate sources (organizations, collectives, and NGOs). El Periódico and Público combine official, corporate, and citizen sources (individuals), although the latter are used to a limited extent. Additionally, these outlets include links to third-party content from various information sources to cover cases uncovered by others, denounce journalistic malpractice, or address local and international issues related to feminism.

The timing of the content is mostly retrospective, with curation of links to materials and sources over a month old and, in some cases, up to 10 years old. Since these are retrospective links, the digital preservation of the documents is particularly important.

The curation techniques employed include summarizing and contextualizing, allowing editors to trace relationships, construct timelines, and define key concepts, as well as analyze structures of violence and inequality and advances in feminist and LGBTBI rights, while also integrating the Ministry of Equality's agenda.

Regarding the morphology of sources, websites predominate, followed by secondary sources and social media, while blogs have minimal presence, reflecting a focus on verified content.

### 4. How to curate gender inequalities and violence

When addressing gender-based violence — including sexist, femicide, sexual, vicarious, institutional, and economic violence — journalists engage in documentary curation, linking and analyzing sources to show how these forms of violence are sustained by social structures. They also craft headlines, provide coverage, and link sources to highlight inequalities in work, health, intimate relationships, caregiving, motherhood, work–family balance, and the political participation and leadership of women and dissenting groups.

Figure 1

(15/02/2024). News article by Anna Requena, gender editor at *ElDiario.es*, on the economic cost of gender based violence in Spain. The headline is based on data from a 2022 study by the Ministry of Equality.



These reports highlight antifeminist discourses circulating from the far-right and the “manosphere” on social media, which seek to delegitimize equality policies by presenting them as an “ideological invention.” Thirty percent of the articles analyze how these discourses are constructed and contested, gaining presence in the media, digital platforms, and institutional politics, especially with the rise of Vox in Spain.

Official sources predominate in the documentation used to cover gender-based violence and inequalities. Institutions such as the Ministry of Equality and its Government Delegation against Gender-Based Violence publish key statistical reports on the [Gender Violence Statical Portal](#).

These reports serve as a reference, and their links are consulted and published when producing information on femicides, complaints, protection orders, and gender-based, sexual, and vicarious violence.

These articles also include curated sources relevant to the prevention and monitoring of gender-based and sexual violence. Among them, a contact email is provided for inquiries and complaints: [0016-online@igualdad.gob.es](mailto:0016-online@igualdad.gob.es). Additionally, links to documents explaining programs for assisting abused women, such as the program [VioPet](#), which provides temporary shelter for the pets of women who are victims of gender based violence.

From these pieces, state level meetings on violence are highlighted, such as the [State Meeting on Vicarious and Gender-Based Violence](#), which facilitate the exchange of experiences and the coordination of public policies on the subject.

Another key official source is the [BOE](#) (Official State Gazette) which allows the public to consult updated regulations on gender equality through the use of its [Digital Legal Library](#). Regulations include reforms such as *Article 709* of the Criminal Procedure Law, which prohibits revictimizing questions to victims of sexual violence, and *Article 681*, which allows the court to bar offensive questions, among other provisions. These articles serve as an example of documentary sources that facilitate the analysis of legal advances and the oversight of the real implementation of these regulations in the courts.

Reports from the Observatory of the [General Council of the Judiciary](#) (CGPJ) are also a valuable documentary source; the analysis of court rulings provides journalists with the opportunity to produce reports on impunity or barriers to accessing justice.

An example of this was the journalistic coverage of the “Dani Alves case.” The former Brazilian footballer was accused in 2023 of raping a young woman in Barcelona, sentenced in 2024 to four and a half years in prison, and ultimately acquitted in 2025 by the High Court of Justice of Catalonia, which determined that there was insufficient evidence. The case sparked significant controversy in Spain.

**Figure 2**

(15/01/2024). Headline produced by Ana Requena based on an official documentary source: a survey by the Center for Sociological Research (CIS) on men's perceptions of equality policies in Spain.

## El 44% de los hombres cree que las políticas de igualdad han ido demasiado lejos y que son ellos los discriminados

La encuesta sobre la percepción de la igualdad publicada por el CIS muestra que las mujeres dedican tres horas al día a las tareas del hogar, frente a las dos de los hombres. La brecha se agranda cuando hay hijos: ellas emplean 6,7 horas; ellos 3,7

— El 44% de los hombres cree que las políticas de igualdad han ido demasiado lejos y que son ellos los discriminados



Foco IGUALDAD



Miles de personas protestan con carteles durante la manifestación encabezada por el Moviment Feminista de Mallorca por el BM. Tomas Moyà / Europa Press

**elDiario.es**  
depende de ti  
Apoya el periodismo independiente desde 3,33 €/mes



Hazte socio/a

Ana Requena Aguilar  
15 de enero de 2024 · 13:14 h Actualizado el 08/03/2024 · 22:55 h 119

SEGUIR AL AUTOR/A



In this process, another key documentary source was the Protocol “[No Callem](#)” from the Barcelona City Council, implemented during Ada Colau’s term to prevent sexual assaults in night-life venues. Its application in this case had international relevance and demonstrated how an official documentary source can guide immediate security and emergency service actions, as well as become news and feature in various headlines.

When covering antifeminist discourses, reports from other official sources are consulted, such as those from the [Center for Sociological Research \(CIS\)](#). Their results are reflected in news articles and reports, such as the survey showing that 44% of men believe equality policies have progressed so much that they now feel discriminated against.

At the international level, notable sources include curated links to the [Instambuul Convention](#), which requires states to prevent and punish violence against women, and to campaigns and resources from [UN Women](#), such as the “[16 Ways to Challenge Rape Culture](#),” which aim to raise awareness and prevent gender-based violence globally. Reports and sessions from international organizations, such as the document from the [74th CEDAW](#) session, serve as a framework for policies and reports on violence.

Alongside official sources, journalists curate links to corporate sources, such as reports from Aspacia, an organization established in Spain in 2020 to coordinate social and equality policies. These organizations, supported in some cases partly with public funds, provide resources such as [violencia-sexual.info](#), guides for psychological, legal, and social support for victims of sexual assault, often linked in gender journalism pieces.

Additionally, journalists incorporate sociological reports — such as surveys conducted by the social research consultancy 40dB for [El País](#) and la [SER](#) — which help contextualize generational gaps regarding feminism and equality, along with studies from organizations such as [Save the Children](#), which warns of the high exposure of young people to digital sexual violence.

Tweets and resources from citizen sources are also curated on X (formerly Twitter) and the web, where expert voices on the subject are highlighted. Among these, the curation of tweets from [Victoria Rosell](#), the former Government Delegate against Gender-Based Violence and a jurist specializing in women’s rights, or [Julia Humet](#), lawyers and political scientists specializing in gender-based violence; as well as documents and reports produced by [Sonia Vaccaro](#), a clinical and forensic psychologist known for her work on gender-based violence and for coining the concept of “vicarious violence.”

In terms of documentary sources from media outlets, journalists link to academic studies that provide evidence on sexuality and justice. For example, [Measuring fe-male genital functions](#) by Roy J. Levin (2004) examines methods for assessing female genital functions. This study helps debunk myths about consent and revictimization. Another study, [Does the sex of judges matter?](#) by Arantegui, Romero, and Tamarit (2024), analyzes how the gender of judges influences sentences for sexual offenses in Spain. The results show that female judges tend to impose harsher penalties.

At the same time, articles from other traditional media outlets are selected and curated to address topics of local or international interest or to complement information, such as headlines highlighting the greater strictness of female judges in sexual offense penalties ([El País](#)), the first law in Castilla-La Mancha against the gender pay gap ([Periódico CLM / Público](#)), the

video *El violador eres tú* (La Nación, Argentine digital outlet), and guidance from the Ministry of the Interior on sexual assaults (20 Minutos).

Apart from gender-based violence, journalists specializing in equality draw on a wide range of documentary sources to cover and craft headlines about the multiple forms of inequality affecting women in various areas: In education and employment, links to official documentary sources are curated, such as the National Institute of Statistics (INE), Eurostat, and the State System of Education Indicators (2023), which provide data on educational segregation, [school dropout rates](#), and [unequal access to employment](#).

Similarly, the Active Population Survey (EPA) and the INE's Wage Structure Surveys are essential for analyzing labour and economic gaps. Likewise, the INE Fertility Survey (2018) provides key insights on motherhood, work-life balance, and delayed childbearing, which are crucial for understanding demographic and labor inequalities.

At the corporate source level, the [GWL Voices](#) report, produced by an organization of nearly 80 women leaders worldwide, served gender journalists in creating various headlines highlighting a key fact: only 67 of the 523 leaders in global organizations over the past 79 years have been women, revealing the glass ceiling in international representation.

In the area of motherhood, journalists draw on resources produced by organizations such as *Yo no renuncio* ([yonorenuncio.com](#)) and *Familias Monomarentales* ([familiasmonomarentales.es](#)) to highlight gender inequalities in work-life balance, single mothers' rights, and caregiving burdens, supporting their articles with reliable data and testimonies. This work of documentation and data analysis — combined with data visualization tools and accessible microdata — strengthens gender journalism, allowing complex figures to be transformed into understandable narratives that highlight persistent inequalities.

To further explore the practice of gender journalism, it is essential to consider how sources and resources produced by feminist movements are selected and organized.

## 5. How to curate sources and resources from feminist movements

News coverage on feminism focuses on the struggles and achievements of feminist movements, with particular emphasis on reproductive and social rights in Spain, and to a lesser extent, in other European and Latin American countries. Internal debates within Spanish and European feminist issues are also addressed, such as the regulation of prostitution, the implications of the so-called "trans law," LGBTIQ+ inclusion, and the interpretation of sexual consent under the "Only Yes Means Yes" law, which legally redefines this concept. Interviews with feminists from cultural, political, and activist spheres are also included, providing perspectives and reflections on the current challenges facing feminist movements.



Figure 3

(04/04/2024). Headline produced by Sabela Rodríguez Álvarez for InfoLibre on the initiative of feminist activists from eight European countries, including Spain, to guarantee the right to free, safe, and legal abortion across the EU.

IGUALDAD

## Abortar en España si en Polonia no se puede: mujeres europeas se organizan para blindar sus derechos

- ▶ Feministas de ocho países lanzan una iniciativa para que la Unión Europea sufrague el desplazamiento entre estados miembros y garantizar la libre interrupción del embarazo
- ▶ "En vez de actuar cuando nos quiten nuestros derechos, la idea es decirles que los vamos a blindar", dice la escritora Cristina Fallarás, una de las impulsoras de 'Mi voz, mi decisión'
- ▶ Exclusivo para socias y socios.

     3 COMENTARIOS



Manifestación en 2022 contra la reforma del aborto en Polonia. EP

 **Sabela Rodríguez Álvarez**  
X@SabelaRuinha

4 de abril de 2024 - 20:57h Actualizado el 05/04/2024 - 06:00h



These pieces are based on a feminist genealogy that contextualizes current struggles and debates within a broader historical framework. Beyond the more punitive feminist perspective or coverage of gender-based violence, they also highlight proactive initiatives that open new positive frameworks.

Figure 4

(02/03/2024). Headline produced by Isabel Valdés for El País on a decade of feminist mobilization in Spain, from the Tren de la Libertad to its influence on social media.



Various sources are compiled, including official sources, [reports](#) from the Ministry of Health on pregnancy terminations, as well as links to European initiatives such as [My Voice, My Choice](#) (ICE), which aims to ensure safe and accessible abortion across the European Union. Nevertheless, significant visibility is given to corporate and citizen sources that support these rights.

Links to materials produced by feminist collectives are highlighted. Examples include: [FemBloc](#), which provides support [resources](#) for women and LGBTQ+ people facing digital violence; studies by the [CATS Association](#), which defends the rights of sex workers and offers social and health services; and publications by [Punt6](#), which promote a more inclusive and equitable feminist urbanism.

There is also extensive publication of articles on feminist movements in Spain highlighting how debates over the prohibition of prostitution and the rights of trans people have divided them. In 2024, for the third consecutive year, the demonstrations were held separately, including in Madrid, with two simultaneous marches:

[El Movimiento Feminista de Madrid](#) called for the prohibition of prostitution and rejected gender self-determination. For its part, the [8M Commission](#) defended trans rights and the eradication of trafficking. It also supported the participation of sex workers in debates that affect them, although it did not take a position on prostitution. In this regard, journalists specializing in equality compiled and curated links to the position papers of both movements to facilitate consultation and analysis.

In terms of curated links to other media, interviews with feminists published in various newspapers are included, along with books from different publishers and external reports.

These links are used to develop a feminist genealogy, that is, to explore how ideas about women and gender have been historically constructed including links to books from feminist publishers such as [La Bella Varsovia](#), [Virus Editorial](#), and [Sudakasa](#), highlighting literature written by women. References to web projects related to historical memory and the visibility of women, such as [Nietas de la Memoria](#), profiles of feminist figures and experts on Wikipedia, and collaborative platforms such as [Mujeres en Red](#), that aim to make women more visible and reduce the gender gap in the digital encyclopedia are collected. Journalistic initiatives that reconstruct the life and work of notable women are included, such as the article by [Vani-tatis](#) in *El Confidencial* about Emilia Pardo Bazán, a writer and advocate for women's rights, or local historical memory reports, such as those produced by *Newtral* on Ana Orantes, a victim of gender-based violence whose case prompted legislative changes in Spain.

Similarly, articles, reports, and interviews from other media that highlight pro-abortion activists are included, such as [Justyna Wydrzynska](#), an activist in Poland, a country with very restrictive abortion laws. Her work was reported by media outlets such as *El País* and later cited by other outlets included in the sample.

These progressive media occasionally include statements from outlets on the conservative spectrum, such as [COPE](#), in order to illustrate positions opposing abortion and to contextualize their origins.

Finally, and to a lesser extent, citizen sources are curated, with links to experts and communicators such as the Twitter profile of [Isabel Mellén](#), a philosopher, art historian, and writer, or the Instagram profile of [Júlia Salander](#), a feminist activist and political scientist; both are feminist communicators.

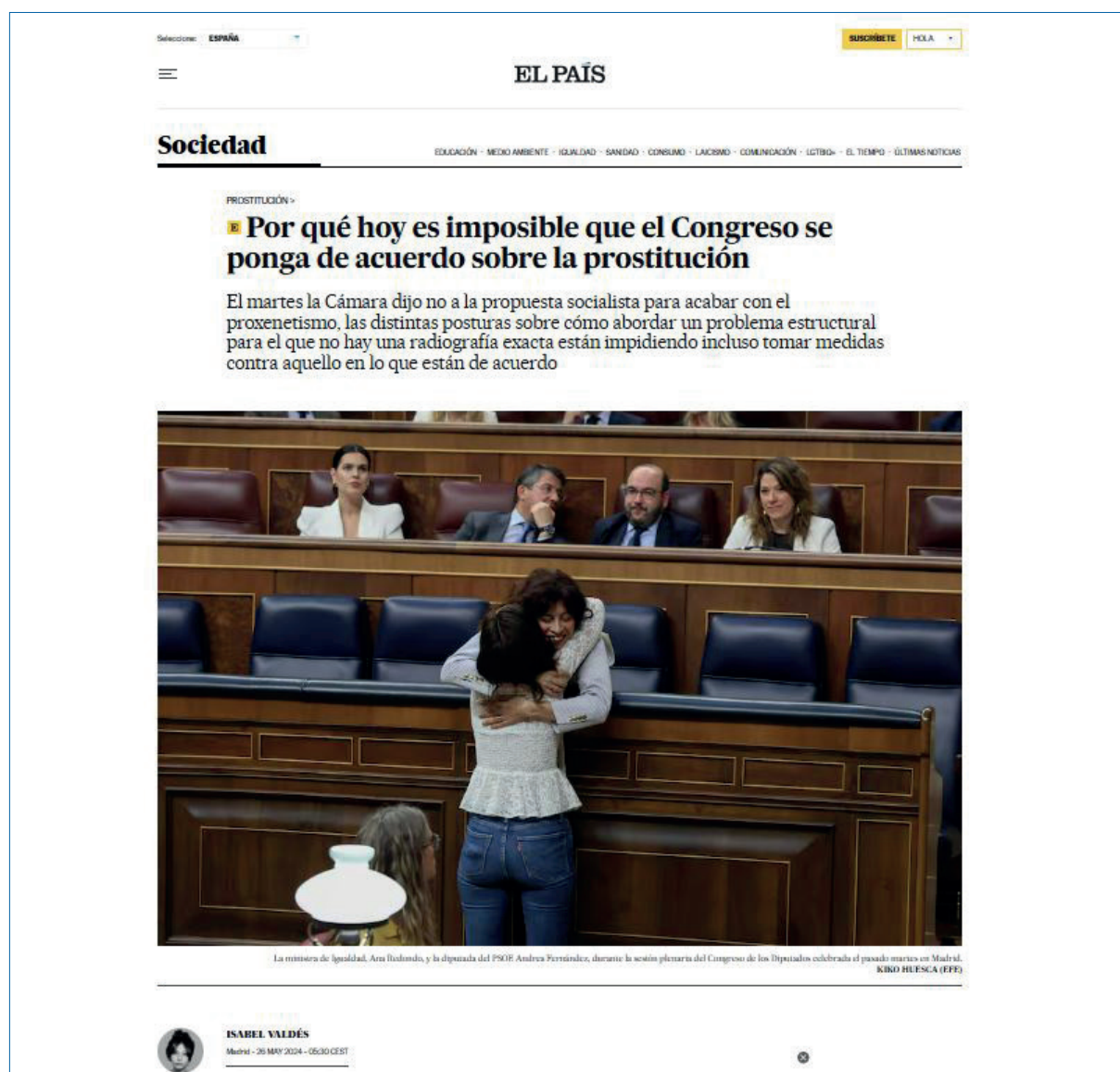
In addition to these voices, curation extends to the regulatory sphere and public policies. These contents are introduced in the following section.

## 6. How to curate equality laws and policies

The articles in this thematic block report on the institutional processing of key laws such as legislation on prostitution, abortion, and the Trans Law, as well as on the activity of the Ministry of Equality and the political stances regarding equality legislation. They analyze how these laws are being implemented and monitor whether they are effectively applied or face delays. They also cover the social and political debates they generate, showing which parties support them, which propose changes, and who oppose them, thus assessing their political and social impact.

Figure 5

Headline produced by Isabel Valdés for El País on the political and feminist divisions that hinder consensus in the Spanish Congress regarding prostitution.



When addressing these topics, mainly official sources are collected, especially documents published in the Press Rooms of political party websites and the Ministry of Equality, as well as the regulations published in the BOE. Among these sources are recent legislative initiatives, such as the proposal of the [PSOE](#) to prohibit pimping (2024) and the [Plan Camino](#) (2022), approved by the Ministry of Irene Montero (Unidas Podemos), which aims to provide economic, labor, and social alternatives to victims of trafficking and sexual exploitation.

Links to documents from corporate sources are also included, such as the letter from [Human Rights Watch](#) to Congress or the publications of [Amnesty International](#), which highlight the impact of punitive laws on the rights of sex workers and contribute human-rights-focused perspectives to media coverage. In particular, those related to sexual legislation, the rights of sex workers, and the reports of [International Planned Parenthood Federation](#), an international organization dedicated to sexual and reproductive health.



Figure 6

(11/06/2024). Headline written by Ana Requena Aguilar for eldiario.es about the autonomous communities that do not provide the emergency contraceptive pill for free, in violation of current law.



In addition to monitoring parliamentary debates on the policies of the Ministry of Equality, the actual implementation of these policies is analyzed through document curation. To this end, links to official sources such as the [BOE](#), but also to corporate sources, such as the report from the [Federación de Planificación Familiar \(SEDRA\)](#), which shows how access to contraception varies across the autonomous communities of Spain, or Mujeres Juristas Themis, which provides reports and legal resources to highlight and combat gender inequality in the legal field. Additionally, it is proposed to include union reports, such as those from [CCOO](#), which analyze gender equality in the workplace and the gender pay gap, as a complement to more comprehensively assess the implementation and impact of equality policies.

Likewise, some citizen and publicly accessible sources are collected and curated, such as the Twitter account of [Elizabeth Duval](#), as well as the Wikipedia entries on [Marina Subirats](#) and [Marisa Soletto](#), which provide relevant contextual and biographical information for the analysis of debates and actors in the field of gender equality. This same logic of source selection and curation is also key for analyzing advances and setbacks in LGBTQI+ rights.

## 7. How to curate advances and setbacks in LGBTQI+ rights

The news in this section deals with the situation of rights and the violence affecting LGBTQI+ people. They report, for example, on setbacks in the care and protection of this community, especially following the change of leadership in the Ministry of Equality.

Figure 7

(09/02/2024). Headline written by Patricia Reguero Ríos for *El Salto* about the difficulties faced by foreign trans people in updating their documents one year after the Trans Law in Spain.



These articles also include links to official sources, such as [Law 4/2023](#), of February 28, for the real and effective equality of trans people and the guarantee of the rights of the LGTBQ+ community. However, there special mention of corporate and citizen sources that defend these rights. For example, editors of media outlets such as *Público* or *El Salto* include links to [petitions from LGTBQ+](#) activists requesting the removal of Isabel García as director of the Institute of Women, due to her statements being considered transphobic and her opposition to the Trans Law. Among corporate sources, links are included to information published by LGTBQ+ organizations, such as the State Federation of Lesbians, Gays, Trans, and Bisexuals (FELGTBI+) on the Ministry of Equality's commitment to the State Pact against hate speech toward vulnerable groups ([felgtbi.org](#)), and the Euforia portal, which provides documentation on the legal recognition of non-binary identities ([euforia.org.es](#)).

Figure 8

(25/01/2024). Headline written by Candela Barro for *Público* about hate campaigns against trans women on Twitter, according to a study by FELGTBI+ and 40dB.



Additionally, citizen sources from the community are given visibility, such as the inclusion of links to interviews with LGBTQ+ activists like Alan Moreno, a Colombian LGBTQ+ activist exiled in Spain after facing threats for being a trans man and for speaking out against sexual harassment and sexism.

Links to other media outside the ecosystem, selected by gender editors, are also curated. For example, among local media, newspapers such as *El Diario de Cantabria* are included, which reports on the complaint by the Student Council of the University of Cantabria regarding messages of LGBTQ-phobia, sexism, and hate directed at university students ([eldiariocantabria.publico.es](http://eldiariocantabria.publico.es)), as well as *El Periódico*, which reports on the increase in LGBTQ-phobic incidents during 2023 ([elperiodico.com](http://elperiodico.com)).

Similarly, international media are linked, such as *The New York Times*, which covers cases of discrimination in U.S. universities, for example, the situation of a non-binary student in Oklahoma ([nytimes.com](http://nytimes.com)) and other related reports ([nytimes.com](http://nytimes.com)).

Information from external media, both local and international, is carefully curated to complement in-house coverage, offer different perspectives, and provide reliable context on situations of discrimination and LGBTQ-phobia. This allows gender editors to provide readers



with a broader and well-documented overview of the issues affecting the LGBTQ+ community in different environments and geographies.

## 8. Importance of the documentary work of gender journalists

To conclude, it is important to highlight that this research shows the scarcity of specialized roles in gender journalism in Spain, with the presence of gender editors limited to three digital newsrooms and only six journalists in managerial positions dedicated to equality sections in national press outlets, reflecting the limited incorporation of these professionals in the media.

The lack of resources and recognition is aggravated by hostility, where gender journalists face harassment, censorship, and questioning of their work, especially from ultraconservative sectors, some digital media, and the so-called “manosphere” on social networks. In addition, mistrust toward official sources and women journalists make their work even more difficult.

Despite this, Spanish journalists continue to carry out rigorous source curation, highlighting equality issues and challenging discourses that deny inequality. This is a form of documentary journalism that demonstrates that gender-based violence remains central to the media agenda, while 30% of the articles produced by gender journalists during the analyzed period address the growth of discourses that deny these inequalities and forms of violence.

As for source curation, during the first half of 2024, more than two thousand documents were curated. Professionals relied primarily on official sources, complemented by corporate, citizen, and media sources. However, a potential ultraconservative government, such as PP or Vox, could limit the production and reliability of these sources, making it necessary to diversify curation in order to guarantee the future continuity of coverage on equality and gender-based violence.

This study analyzes the current practices of gender journalism, pointing out that, although progress has been consolidated, these are situational dynamics and future challenges persist. Among them is the need to strengthen alternative documentary networks that ensure critical, rights-based journalism even in adverse scenarios, as well as the creation of safe spaces to work and support networks with colleagues, lawyers, and health professionals, fostering alliances and rejecting self-censorship.

Although all the selected media outlets curate a diversity of documentary sources, variations in their approaches are observed: while journalists at El País and Infolibre prioritize official sources on violence, inequalities, and public policies, the editor at El Salto emphasizes proactive, community, and feminist perspectives, whereas professionals at El Periódico and Público combine official, corporate, and citizen sources, though with limited use of the latter.

Content curation focuses on retrospective links, some dating back as far as ten years, which underscores the importance of digital preservation. This practice ensures access to verifiable sources and allows for maintaining the continuity and integrity of analyses on inequalities and denialist discourses. However, a recurring challenge is the poor preservation of digital documents, as over time many links break, making verification and the reconstruction of debates difficult.

These documentary sources are linked through techniques of summarization and contextualization, which facilitate the construction of chronologies, the definition of key concepts, and the analysis of structures of violence, inequality, and advances in feminist and LGBTQ+ rights, while also incorporating the agenda of the Ministry of Equality.

Likewise, monitoring the Ministry's activity, the implementation of its policies and regulations, and the proper management of allocated funds — at the state, regional, and municipal levels — together with tracking the progress and challenges in rights and violence affecting the LGTBQ+ community, contributes to institutional accountability. This makes it possible to highlight achievements, point out shortcomings, and demand responsibility in the promotion of equality.

Regarding the morphology of these sources, websites predominate, followed by secondary sources and social networks, while blogs have minimal presence, reflecting an emphasis on verified content.

Although all the media outlets analyzed target progressive audiences, each editor develops a distinct approach to feminism and gender inequality. While *Público* and *El Salto* compile more materials on trans rights, migrants, and the participation of sex workers, other outlets also include debates on the abolition of prostitution and opposition to gender self-determination, while maintaining, in all cases, a pro-rights stance and respect for the diversity of feminist currents.

This source-curation journalism fulfills informative, educational, and strategic functions, providing tools and arguments to those who promote equality. Although it does not seek to persuade denialists, it strengthens rights-based narratives and supports policies, content, and critical reflections. In this way, journalists position the equality agenda in the public sphere, consolidating coverage that highlights violence and inequalities while also proposing solutions and new frameworks for thought.

From this research, new lines of study emerge, notably the analysis of themes and source curation by journalists across different periods, countries, or political contexts, which would enrich comparison and lead to a better understanding of the global dynamics of gender journalism.

It would also be relevant to examine the curation and use of documentary sources in media lacking specialized gender-equality roles, as well as in those closer to far-right positions, to understand how such conditions affect coverage, source selection, and topic treatment. It should be noted that this research has focused exclusively on content published by six journalists holding editorial leadership positions; therefore, it would be valuable to expand the study to other works produced by professionals in equality sections and, furthermore, to interview them in order to explore their topic selection and source-curation practices in greater depth, thereby complementing the information presented in this chapter

## 9. Funding

This work is part of the Project “Parameters and strategies to increase the relevance of media and digital communication in society: curation, visualisation and visibility (CUVICOM)”. Grant PID2021-123579OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU.

## 10. References

Amnistía Internacional. (2024). *Muteadas: el impacto de la violencia digital contra las periodistas*. Amnistía Internacional Argentina. [https://amnistia.org.ar/wp-content/uploads/de-lightful-downloads/2024/10/Muteadas\\_InformeCompleto.pdf](https://amnistia.org.ar/wp-content/uploads/de-lightful-downloads/2024/10/Muteadas_InformeCompleto.pdf)

- Arantegui, L., Romero, P., & Tamarit, J. M. (2024). Does the sex of judges matter? Analysis of sentencing in sex crimes in Spain. *Criminology & Criminal Justice*, First published online. <https://doi.org/10.1177/17488958241239785>
- Bailo, F., Meese, J., & Hurcombe, E. (2021). The institutional impacts of algorithmic distribution: Facebook and the Australian news media. *Social Media + Society*, 7(2). <https://doi.org/10.1177/20563051211024963>
- Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, 33(2), 122–139. <https://doi.org/10.1177/0267323118760317>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Centro de Investigaciones Sociológicas. (2024, 15 de enero). Avance de resultados del estudio 3428: Percepciones sobre la igualdad entre hombres y mujeres y estereotipos de género. <https://www.cis.es/es/-/avance-de-resultados-del-estudio-3428-percepciones-sobre-la-igualdad-entre-hombres-y-mujeres-y-estereotipos-de-genero>
- Cui, X., & Liu, Y. (2017). How does online news curate linked sources? A content analysis of three online news media. *Journalism*, 18(7), 852–870. <https://doi.org/10.1177/1464884916663621>
- Europa Press. (2024, 4 de septiembre). Plataforma Trans insta a Redondo a desarrollar instrucciones y protocolos de la Ley Trans para “hacerla efectiva”. <https://www.europapress.es/sociedad/noticia-plataforma-trans-insta-redondo-desarrollar-instrucciones-protocolos-ley-trans-hacerla-efectiva-20240904131041.html>
- European Institute for Gender Equality. (2023). *Gender Equality Index 2023*. Publications Office of the European Union. <https://op.europa.eu/en/publication-detail/-/publication/73a85107-9b0b-11ee-b164-01aa75ed71a1>
- Feezell, J. T. (2017). Agenda setting through social media: The importance of incidental news exposure and social filtering in the digital era. *Political Research Quarterly*, 71(2), 482–494. <https://doi.org/10.1177/1065912917744895>
- Fletcher, R., & Nielsen, R. K. (2017). Are people incidentally exposed to news on social media? A comparative analysis. *New Media & Society*, 20(7), 2450–2468. <https://doi.org/10.1177/1461444817724170>
- García-García, J. (2025). Respuestas normativas para enfrentar el discurso del odio desde la experiencia europea. *European Public & Social Innovation Review*, 10, 1–12. <https://doi.org/10.31637/epsir-2025-1904>
- García-Mingo, E., & Díaz Fernández, S. (2023). Wounded men of feminism: Exploring regimes of male victimhood in the Spanish manosphere. *European Journal of Cultural Studies*, 26(6), 897–915. <https://doi.org/10.1177/13675494221140586>
- Guallar, J. (2014). Content curation en periodismo (y en documentación periodística). *Hipertext.net*, (12). <https://raco.cat/index.php/Hipertext/article/view/275781>
- Guallar, J., & Codina, L. (2018). Journalistic content curation and news librarianship: Differential characteristics and necessary convergence. *El Profesional de la Información*, 27(4), 778–791. <https://doi.org/10.3145/epi.2018.jul.07>

- Guallar, J., Franch, P., Cascón-Katchadourian, J., & Boté-Vericad, J.-J. (2025). The rise of curated newsletters in media: A case study of the New York Times. *Journalism Practice*. <https://doi.org/10.1080/17512786.2025.245733>
- Guallar, J., Pedraza, R., Pérez-Montoro, M., & Antón, L. (2021). Curación de contenidos en periodismo. Indicadores y buenas prácticas. *Revista Española de Documentación Científica*, 44(2). <https://doi.org/10.3989/redc.2021.2.1742>
- Guerrini, F. (2013). *Newsroom curators & independent storytellers: Content curation as a new form of journalism*. Reuters Institute.
- Iranzo-Cabrera, M., Figueras-Maz, M., & Mauri-Ríos, M. (2023). Journalistic self-regulation for equality: The role of gender editing in Spain. *Journal of Media Ethics*, 38(1), 2–15. <https://doi.org/10.1080/23736992.2022.2158336>
- Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). SAGE Publications.
- Levin, R. J. (2004). Measuring female genital functions—a research essential but still a clinical luxury? *Sexual and Relationship Therapy*, 19(2), 191–200. <https://doi.org/10.1080/14681990410001691406>
- Lopezosa, C., Guallar, J., Codina, L., & Pérez-Montoro, M. (2023). Content curation and journalism: exploratory scoping review and expert opinion. *Revista Mediterránea de Comunicación*, 14(1), 205–223. <https://doi.org/10.14198/MEDCOM.22810>
- Ministerio de Igualdad. (2023, 21 de noviembre). *Ana Redondo toma posesión como ministra de Igualdad*. <https://www.igualdad.gob.es/home/ana-redondo-toma-posesion-como-ministra-de-igualdad/>
- ONU Mujeres. (2025). *Los derechos de las mujeres bajo examen: 30 años después de Beijing*. <https://www.unwomen.org/es/digital-library/progress-of-the-worlds-women>
- Pan, R., García-Díaz, J. A., & Valencia-García, R. (2025). Spanish MTLHateCorpus 2023: Multi-task learning for hate speech detection to identify speech type, target, target group and intensity. *Computer Standards & Interfaces*, 94, 103990. <https://doi.org/10.1016/j.csi.2025.103990>
- Reuning, K., Whitesell, A., & Hannah, A. L. (2022). Facebook algorithm changes may have amplified local Republican parties. *Research & Politics*, 9(2). <https://doi.org/10.1177/20531680221103809>
- Spinetta, A. B. (2022). Editoras de género en Argentina. Avances y desafíos de un rol en construcción. *BiD: Textos Universitaris de Biblioteconomia i Documentació*, (49). <https://doi.org/10.1344/BiD2022.49.09>
- UNESCO. (2019). *The Chilling: Tendencias mundiales en la violencia en línea contra mujeres periodistas*. <https://www.unesco.org/es/articles/the-chilling-tendencias-mundiales-en-la-violencia-en-linea-contra-mujeres-periodistas>
- Virgili, J. R., & Puche, J. S. (2019). Medios de comunicación y opinión pública en España: Una aproximación desde la teoría de agenda setting. *Panorama Social*, 30, 27–39. <https://dialnet.unirioja.es/servlet/articulo?codigo=7207334>
- Waisbord, S. (2018). Truth is what happens to news. *Journalism Studies*, 19(13), 1866–1878. <https://doi.org/10.1080/1461670x.2018.1492881>

# SEOVisFrame: best practices and method for evaluating the SEO optimization of storytelling data visualisations

**Rubén Alcaraz-Martínez**

Universitat de Barcelona, Spain

<https://orcid.org/0000-0002-7185-0227>

**Mari Vállez**

Universitat de Barcelona, Spain

<https://orcid.org/0000-0002-3284-2590>

**Raquel Escandell-Poveda**

Universidad de Alicante, Spain

<https://orcid.org/0000-0002-8398-1873>

**Diana Bajaña-Cedeño**

Universitat de Barcelona, Spain

<https://orcid.org/0009-0008-7395-3028>

---

Alcaraz-Martínez, R., Vállez, M., Escandell-Poveda, R., & Bajaña-Cedeño, D. (2025). SEOVisFrame: best practices and method for evaluating the SEO optimization of storytelling data visualisations. In J. Guallar, M. Vállez, & A. Ventura-Cisquella (Coords). *Digital communication. Trends and good practices* (pp. 69-84). Ediciones Profesionales de la Información. <https://doi.org/10.3145/cuvicom.06.eng>

---

## Abstract

Storytelling data visualisations have gained prominence in digital media, enabling the communication of complex information in an accessible and engaging manner. However, their impact and reach on search engines largely depend on their visibility in search results. Based on this premise, this chapter examines the key ranking factors for such information products and proposes a set of best practices and a methodology to evaluate its SEO optimisation through a list of indicators, referred to as SEOVisFrame.

## Keywords

Storytelling data visualizations; Data-driven journalism; Media outlets; Search Engine Optimisation (SEO); Search engine visibility.

---

## 1. Introduction

In the information age, data-driven journalism has emerged as a powerful instrument for structuring, analysing, and telling stories based on large volumes of data. This journalistic speciality, which combines data analysis techniques with visual narratives, not only helps to unravel the complexity of data but also makes it accessible and understandable to the public (Córdoba-Cabús, 2020). Its evolution has been driven by the rise of digitisation and the availability of open data, marking a transformation in the practice of traditional journalism that is increasingly driven by data (Radcliffe & Lewis, 2019).

Although its origins date back to the second half of the 20th century, the modern practice was consolidated in the late 2000s and early 2010s (Ferreras Rodríguez, 2015), driven by the proliferation of available open data, particularly from public administrations, the emergence of new technological tools focused on the analysis of large datasets and the creation of interactive visualisations, alongside a media context in which audiences demanded innovative new narratives. Some of its advances are the result of the investigation into the so-called *Panama Papers* of the International Consortium of Investigative Journalists (2016), or applications stemming from this approach, such as those published by the Civio Foundation (*Dónde van mis impuestos or the Indultómetro*) (Ferreras Rodríguez, 2015).

The integration of departments or units specialised in data journalism within cybermedia became a reality during the 2010s (Peiró & Guallar, 2013). During the COVID-19 pandemic, data journalism played a crucial role, becoming an essential intermediary between institutions and society. The abundance of statistics and their lack of consistency highlighted the importance of this approach in organising, interpreting, and communicating relevant information (Córdoba-Cabús et al., 2020).

Alexander and Vetere (2011) propose a list of parameters to ensure the quality of content in data journalism, such as trust in the data presented, the integration of the surprise factor, transparency in methodological processes, the identification of a leader, the use of accessible language, and the possibility of access to the original data. The work of Córdoba-Cabús (2020) analyses some of the most internationally recognised data visualisations emerging from the journalistic sector, with the aim of synthesising a list of common elements in this journalistic practice. Its variables for analysis include subject matter, narrative style, story types,



dimension, elements of focus, purpose, method of communication, source, type of analysis performed, types and ratio of visualisations. The quality and effectiveness of these narrative products depend on parameters such as transparency, and the selection and treatment of sources, as well as the appropriateness and innovation of the forms selected to present the information (Córdoba-Cabús, 2020).

Since the late 1990s and early 2000s, when both major international and national newspapers introduced their respective digital versions, search engine optimisation (SEO) has become an indispensable tool for journalism, as it directly influences the visibility and accessibility of news content. Given that search engines are the primary access point for finding information on the web, the capacity of journalistic content to rank high in search results is crucial for reaching a wider readership. SEO practices not only enhance the discoverability of articles but also ensure that the content is presented in a way that meets the evolving demands of both algorithms and users. Given the competitive nature of the online media landscape, effective SEO strategies are vital for news organisations seeking to maintain relevance and engage audiences in an increasingly crowded digital environment.

The number of specific SEO ranking factors and their relative weight in search engine algorithms remain largely unknown, posing a daily challenge for professionals in the field. This tacit knowledge has been primarily disseminated through a range of specialised books (Maciá Domene, 2019; Enge et al., 2023; Lewandowski, 2023; Vicente et al., 2024), as well as through less academic channels (Semrush, 2024; Dean, 2024) and, to a lesser extent, scientific papers (Chotikitpat et al., 2015; García-Carretero et al., 2016; Lopezosa et al., 2018; Ziakis et al., 2019; Almukhtar et al., 2021), some of which are specifically focused on the online media (Dick, 2011; Giomelakis & Veglis, 2015; Lopezosa et al., 2019; 2020; Giomelakis, 2023). In addition to these sources, the technical documentation published by search engines is crucial, with Google being the most prolific. This documentation covers topics ranging from general search engine fundamentals (Google, 2024a) to more specific aspects, such as site optimisation aligned with Google's EEAT (Experience, Expertise, Authority, and Trust) guidelines (Google, 2024b) or SEO-focused recommendations for image optimisation (Google, 2024c).

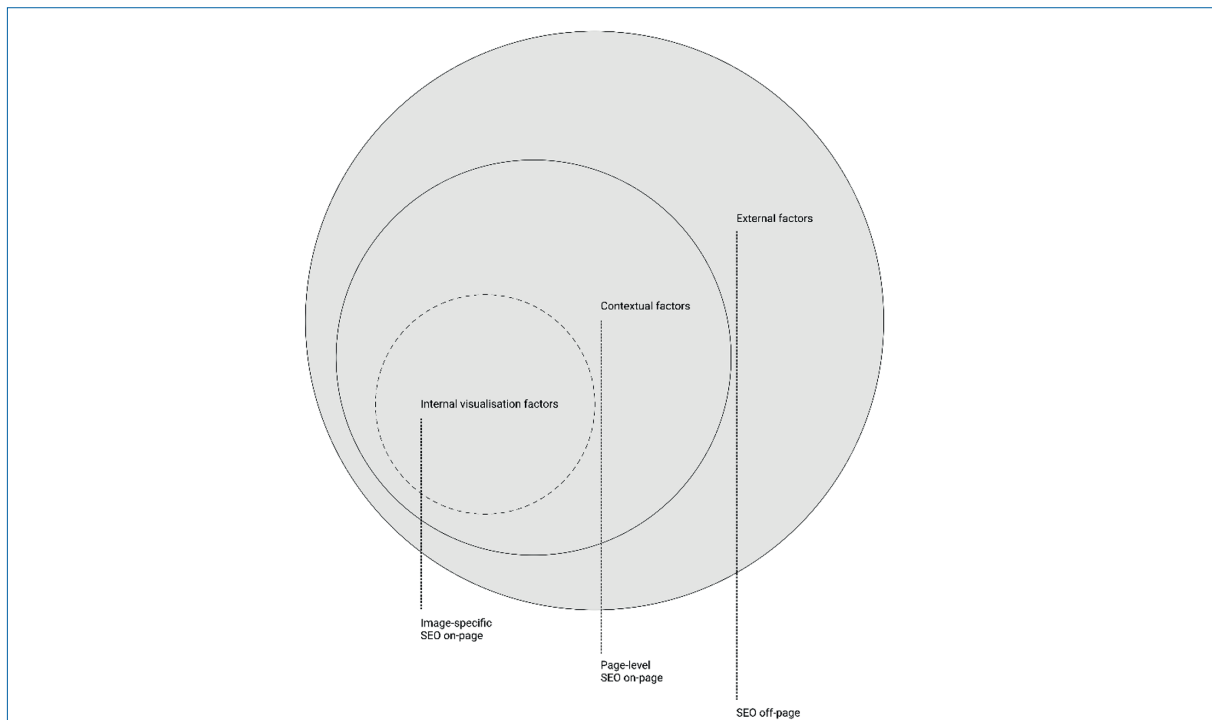
The articles published under the umbrella of data journalism are complex products that incorporate multiple forms of data visualisation, such as tables, interactive statistical charts, maps, infographics, and rich media animations, accompanied by narratives that typically form a storyline (Domínguez, 2016). This approach has led to formats such as long-form journalism and techniques like scrollytelling (Seyser & Zeiller, 2018). Therefore, addressing the SEO optimisation of such products requires the analysis of various dimensions (textual and graphic elements), each of which is linked to different ranking factors, as well as their visibility in different search engines (e.g. Google Search and Google Images).

The most common way to classify SEO factors is based on the degree of control and influence that can be exerted over them. According to this classification, a distinction can be made between on-page and off-page SEO ranking factors (Lewandowski, 2023). On-page SEO can be further divided into technical SEO and content SEO (Escandell-Poveda et al., 2021), two fundamental yet distinct areas that impact a website's performance and visibility.



**Figure 1**

*Classification of ranking factors for storytelling data visualisations.*



Technical SEO focuses on optimising ranking factors related to page crawling and indexing, mobile compatibility, redirect management, and overall site performance and loading speed (Krstić, 2019; Vicente et al., 2024).

Moreover, the use of the HTTPS protocol has been a ranking factor since 2014 (Google, 2014). When implementing HTTPS on a domain, it is important to ensure that all pages, as well as the resources required to render their content, such as images and other elements, are loaded using this protocol. Google (2014) recommends using relative URLs for resources that reside on the same secure domain to ensure the correct protocol. This can also improve performance if absolute URLs are misconfigured (e.g., using HTTP instead of HTTPS). In such cases, browsers can resolve relative URLs more quickly, as they avoid potential redirects.

The integration of Schema.org structured data has been another key technical SEO element on many websites since its introduction by Google in 2011. Structured data is a standardized format for providing information about a page and its content. Structured data is implemented using in-page markup on the page to which the information applies. The inclusion of structured data markup enhances news indexing and improves the accuracy of search results. Ambiguity in content interpretation by search algorithms is also reduced (Salem et al., 2025). When a news article page has structured data, Google can also use that information to display a rich snippet for the piece. Although literature findings indicate that ranking remains the most critical factor in perceived relevance, in certain cases, the richness of snippets can capture user attention (Marcos et al., 2015). In other words, these elements can significantly influence users' behaviour and decision-making (Rodas et al., 2016). Structured data must be representative of the main content of the page, syntactically correct, ensure that the referenced content is visible to users, and comply with the guidelines for each specific structured data feature (Google, 2025a). In this case, *NewsArticle* structured data is the most

relevant. However, structured data for images can also be useful, as it enables Google Images to display additional details such as the creator, rights, or credit information.

Good page performance is also crucial with websites increasingly being accessed via mobile devices. To assess this, Google introduced the Core Web Vitals (CWV) in 2020, incorporating them into its algorithm as a ranking factor in 2021. These are a set of metrics designed to measure the performance of web pages (Maciá Domene, 2019; Alcaraz Martínez, 2022; Vicente et al., 2024), including the LCP (Largest Contentful Paint), INP (Interaction to Next Paint), and CLS (Cumulative Layout Shift). The LCP, which measures the load time of the largest visible text or image element in the viewport, is an indicator of perceived load speed. INP measures the responsiveness of pages to user interactions (such as clicks or keystrokes). Finally, CLS is a measure of visual stability and describes unexpected layouts shifts in the viewport during page loading. Storytelling through data visualisation presents significant CWV challenges, as it often integrates multiple complex images on a single page. The optimisation of these files (selection of the appropriate bitmap format, compression, dimensions, etc.) or the creation of crawlable, search engine-friendly vector images can be challenging.

On the other hand, content SEO aims to meet users' needs by creating texts and other types of content (such as images, multimedia and interactive content) optimised for relevance based on keywords and the search intent expressed by users (Alcaraz Martínez, 2022).

The most important on-page content ranking factors for any page are equally significant for the storytelling data visualizations under study. The inclusion of the keyword in the title meta tag, meta description, URL, h1-h3 headings, or within the body of the page itself with an appropriate density, are fundamental factors (Chotikitpat et al., 2015; Almukhtar et al., 2021; Maciá Domene, 2019; Lewandowski, 2023; Dean, 2024; Google, 2024a; Semrush, 2024).

Although no longer as important as it once was, incorporating keywords into the <title> tag remains an important on-page SEO signal. While meta descriptions were historically an important ranking factor, their influence has diminished. In recent years, Google has increasingly generated snippets directly from the page content. However, Google may still use the meta description HTML element if it provides users with a more accurate summary of the page than content extracted directly from the site.

All pages should include an <h1> heading tag with content closely aligned to the <title> meta tag. Unlike the <title> meta tag, the <h1> tag does not contribute to the generation of SERP snippets. Consequently, strategies such as incorporating the branded keyword (e.g., the name of the media outlet) are commonly applied in meta titles but not in top-level headings. Optimising the <h1> tag for SEO involves crafting story titles that align with the search terms for which visibility is sought. This does not preclude creativity, as media virality and news trends can sometimes render titles such as *The Panama papers* highly effective.

Within the <main> element of a page, multiple levels of headings can be included, ranging from <h2> to <h6>. From an SEO perspective, those higher in the hierarchy, particularly the <h2> and <h3> tags, carry greater relevance. According to Mueller (2015), coordinator of Google Search Relations, "heading tags in HTML help us to understand the structure of the page." While a single page may aim to rank for multiple keywords, structuring the content into distinct thematic sections, each with its own specific headings, is a widely adopted SEO practice (Maciá Domene, 2019).

Images must be optimised for accessibility to crawlers and be easily indexed. Bitmap images should be provided via HTML elements such as `<img>`, `<figure>` or `<picture>`, but never as CSS background images, which is quite common when implementing Parallax effects typical in storytelling. In terms of relevance, keywords should be included in the alternative text, file name, image caption, and in the contextual content (Chotikitpat et al., 2015; Maciá Domene, 2019; Lewandowski, 2023; Semrush, 2024). The alternative text is the most important attribute for providing additional details about an image (Google, 2024c), while also improving accessibility for users who cannot see images on web pages. The primary keyword of the page does not need to be added to all images, and excessive keyword usage (keyword stuffing) should be avoided. Selecting appropriate images to illustrate content is a crucial first step. When done effectively, all images will align with the article's subject matter, reinforcing its overall relevance and enabling the creation of natural, well-optimised alternative texts that serve both SEO and accessibility purposes (Alcaraz Martínez, 2024). Google recommends using descriptive filenames rather than generic ones, even localising them according to the language of the content (Google, 2024c). Although the exact impact of this factor remains unknown, its status as a recommended practice suggests it is beneficial. Images should be placed near relevant text and on pages that align with their subject matter (Lopezosa et al., 2018). While no official guidelines address the proximity of image-related content, logic suggests considering the paragraphs immediately before and after each image. Another important element, given its direct semantic relationship with the image, is the `<figcaption>` element. The `<figcaption>` element within a `<figure>` grouping associates one or more images with a visible description (image caption) as follows:

**Figure 2**

Example of the use of the `<figcaption>` tag.

```
<figure>
  
  <figcaption>Figure 1: Figcaption.</figcaption>
</figure>
```

Metadata is best to be included in IPTC format, a technology that Google uses to extract certain data from images, as it can improve their visibility in Google Images (Google, 2025b). IPTC metadata is embedded directly into the image, allowing both the image and its metadata to remain intact when used across different pages. Unlike structured data, IPTC metadata only needs to be added once per image, regardless of how many pages it appears on. The properties extracted by Google include *Copyright Notice*, *Creator*, *Credit Line*, *Digital Source Type*, *Licensor URL*, and *Web Statement of Rights*.

According to Google (2024c) "high-quality photos appeal to users more than blurry, unclear images". Also, sharp images are more appealing to users in the result thumbnail and can increase the likelihood of getting traffic from users. It is important to strike a balance between quality and weight in order to meet the CWV metrics discussed above and to satisfy user expectations and user experience.

In 2020, Google introduced the `max-image-preview` robots meta tag, allowing publishers greater control over how images from their sites appear in Google Discover. When the `max-image-preview:large` meta tag is added to the header of a webpage, it signals that Google may display the site's images in larger formats. This feature enhances the visual presentation of content on Search surfaces, such as Discover, creating a more engaging user experience.

SVG images are embedded in an HTML page in two main ways: either as an external file referenced by the `<img>` element or by directly incorporating the SVG code within the HTML document. The first method is no different from the practice of using bitmap images within the same element and requires the inclusion of alternative text, along with the other elements mentioned above. Embedding SVG code within the page offers several advantages such as the ability to manipulate SVG elements using scripts and to modify their appearance via CSS. Additionally, certain SVG elements and attributes contribute to accessibility, such as `<desc>`, `<title>`, and `<text>`, which provide textual descriptions for visually impaired users. These elements can also be indexed by search engines (Ferraz, 2017). The `<desc>` and `<title>` elements are specifically designed to describe images. While they are not visually rendered on the page, the `<title>` element may appear as a tooltip when hovered over, depending on the browser. The `<text>` element defines a graphical text component within the SVG. It can include attributes that control visual properties such as text direction, positioning, and fill. Unlike `<desc>` and `<title>`, `<text>` is rendered visibly on the page, and its placement depends both on its position in the code and on the X and Y coordinates defined within the SVG container.

**Figure 3**

SVG vector bar chart example with XML markup, including `<title>`, `<description>`, and `<text>` elements.

```
<svg width="450" height="300" viewBox="0 0 450 300" role="img" aria-labelledby="chartTitle chartDesc">
  <title id="chartTitle">Children at risk of poverty or social exclusion in Spain</title>
  <desc id="chartDesc">This bar chart displays the percentage of children at risk of poverty or social exclusion in Spain from 2019 to 2022. The x-axis represents the years, and the y-axis represents the percentage.</desc>
  <text x="225" y="290">Year</text>
  <text x="20" y="150">Percentage (%)</text>
  <rect class="bar" x="70" y="120" width="60" height="150"></rect>
  <rect class="bar" x="160" y="100" width="60" height="170"></rect>
  <rect class="bar" x="250" y="130" width="60" height="140"></rect>
  <rect class="bar" x="340" y="110" width="60" height="160"></rect>
  <text class="value-label" x="100" y="115">33.0%</text>
  <text class="value-label" x="190" y="95">35.8%</text>
  <text class="value-label" x="280" y="125">31.1%</text>
  <text class="value-label" x="370" y="105">34.5%</text>
  <text x="100" y="280">2019</text>
  <text x="190" y="280">2020</text>
  <text x="280" y="280">2021</text>
  <text x="370" y="280">2022</text>
</svg>
```

In the previous example, the `aria-labelledby` attribute of the Google-recommended (2024c) W3C WAI-ARIA specification, is also included. This allows an element, in this case the visualisation through its `<svg>` tag, to be related to a textual description. The value of the `aria-labelledby` attribute refers to the HTML element (in this case, `<desc>` containing the description).

Originally published in 2013 and periodically updated (Google, 2024b), the *Search quality rating guidelines* serve as the reference document for understanding what Google considers when assessing the quality of web page content, based on the criteria used by its reviewers. These guidelines introduce the concept of EEAT, which is not only linked to factors such as a page's ability to satisfy search intent or its compatibility with mobile devices but also to various elements associated with the experience, expertise, authority, and trust conveyed by a site through its content. This is particularly relevant for media and sensitive-category websites, including those related to economic or health issues. The direct experience of the author and their familiarity with the subject they write about constitutes the first key factor. Being a prestigious media outlet associated with data or investigative journalism is one way to fulfil this criterion. Expertise refers to the level of specialisation of both the author and the publication in a specific subject, reflecting the accumulated experience that qualifies them as specialists. Authority, in this context, pertains to the reputation of the media outlet (and



its website) as well as that of the content author. This authority is assessed based on factors such as the search volume associated with branded keywords (related to the media outlet and author), inbound links, and social mentions. Finally, trustworthiness determines the reliability, honesty, and accuracy of the published information. This is evaluated through technical aspects such as the use of HTTPS, the presence of legal information pages about the media outlet, dedicated author pages featuring a biography relevant to their area of expertise, and links to their published content and professional social media profiles. In the context of data journalism, it is also essential to provide information on the origin or source of the data, as well as the methodology employed.

**Figure 4**

*El Nacional.cat* has a section (*/firmas/*) dedicated to all its editors. Each article links to its authors, whose profiles include a biography, social media links, content tags, and related authors.



Finally, off-page ranking factors are factors that cannot be managed through the site's content or technical configuration but that influence its authority and ranking. These factors include the number of inbound links (Chotikitpat et al., 2015; Lopezosa et al., 2019), certain authority metrics applicable at the domain or page level (domain authority and page authority) (García Carretero et al., 2016), traffic volume (Krstić, 2019), and others generated in response to user interaction (such as click-through rate, dwell time, bounce rate or pogo sticking) (Maciá Domene, 2019).

It is imperative to consider the influence of external links and the diversity of referring domains when evaluating a page's ranking potential, given their impact on search engine algorithms. These factors serve as indicators of authority and relevance, contributing significantly to visibility in search results. Domain authority, page authority, and estimated monthly traffic have been identified as key benchmarks that help to explain why some pages achieve strong rankings despite not excelling in certain aspects of on-page optimisation. While metrics such

as domain and page authority are not direct ranking factors used by search engines, they offer valuable insights into a site's potential for optimal performance. High estimated traffic can be indicative of a page's capacity to attract users, which may in turn correlate with good rankings. When comparing the positioning of two journalistic resources competing for the same keywords, these metrics can help explain the greater ranking potential of certain content or media over others.

## 2. Proposal

SEOVisFrame is a proposal based on a set of indicators for evaluating narrative data visualisations. These indicators encompass SEO ranking factors as well as other elements that can indirectly enhance the visibility and ranking potential of such articles. They may be understood both as a set of best practices and as criteria for assessing existing content.

A total of 34 factors were identified, and are listed in Table 1. Each factor is assessed using a three-point qualitative Likert scale, coded from 0 to 2, to facilitate the calculation of average scores and enable comparisons.

**Table 1**  
*Indicators for the analysis of SEO optimisation of data visualisations.*

ID	Factor	Description	Scope of applicability	Analytical method / Rating scores
1	Keyword in title meta tag	The keyword is included in the <title> meta tag.	All	Manual <title> evaluation. <b>Rating scores</b> 0: The page does not have a title, or the keywords are not included. 1: The title is of partial relevance. 2: The title is fully relevant.
2	Keyword in meta description	The keyword is included in the meta description.	All	Manual meta description evaluation. <b>Rating scores</b> 0: The page does not have a meta description, or the keywords are not included. 1: The meta description is of partial relevance. 2: The meta description is fully relevant.
3	Keyword in URL	The keyword is included in the URL.	All	Manual URL evaluation. <b>Rating scores</b> 0: The keywords are not included in the URL. 1: The URL is of partial relevance. 2: The URL is fully relevant.
4	Keyword in <h1> tag	The keyword is included in the <h1> tag.	All	Manual <h1> tag evaluation. <b>Rating scores</b> 0: The page does not have an h1 tag, or the keywords are not included. 1: The <h1> tag is of partial relevance. 2: The <h1> tag is fully relevant.

ID	Factor	Description	Scope of applicability	Analytical method / Rating scores
5	Keyword in <h2> or <h3> tags	The keyword is included in the <h2> / <h3> tags.	All	Manual <h2> and <h3> tags evaluation. <b>Rating scores</b> 0: The page does not have an h2 or h3 tags, or the keywords are not included. 1: The <h2> / <h3> tags are partial relevant. 2: The <h2> / <h3> tags are fully relevant.
6	Latent semantic indexing keywords in content (LSI)	Latent semantic indexing keywords (synonyms, semantically related terms, etc.) are incorporated into the page content.	All	Manual content evaluation <b>Rating scores</b> 0: Keywords conceptually related to target keywords are not included. 1: Some keywords conceptually related to target keywords are included. 2: Many keywords conceptually related to target keywords are included.
7	HTTPS usage	The page and its resources use the HTTPS protocol.	All	Manual URL evaluation <b>Rating scores</b> 0: Does not use https. 1: Partially uses https. 2: Use https.
8	Schema.org markup	The page includes appropriate and valid Schema.org markup.	All	Manual Schema.org code evaluation. <b>Rating scores</b> 0: Does not use schema.org. 1: Schema.org is used, but the type implemented is either inappropriate or contains code errors. 2: Use an appropriate and valid schema.org type.
9	Data source	The article references the data source.	All	Manual content evaluation <b>Rating scores</b> 0: Does not provide the data source. 1: Provides the data source.
10	Link to data source	The article includes a link to the data source.	All	Manual content evaluation <b>Rating scores</b> 0: Does not provide the data source link. 1: Provides the data source link.
11	Authorship	The article references its authors.	All	Manual content evaluation <b>Rating scores</b> 0: Does not provide the authorship. 1: Provides the authorship.
12	Link to the author's bio or content page	The article includes links to author bio or information pages.	All	Manual content evaluation <b>Rating scores</b> 0: Does not provide the author' bio or info page link. 1: Provides the author' bio or info page link.
13	Methodology	The article includes details about the research methodology used.	All	Manual content evaluation <b>Rating scores</b> 0: Does not provide information about the methodology. 1: Provides information about the methodology.



ID	Factor	Description	Scope of applicability	Analytical method / Rating scores
14	Compatible technologies	Data visualizations and accompanying text are implemented using technologies that are easy to crawl and index.	All	Manual source code evaluation <b>Rating scores</b> 0: The technologies used are not compatible with search engines' crawling and indexing processes. 1: The technologies used are partially compatible with search engines' crawling and indexing processes. 2: The technologies used are compatible with search engines' crawling and indexing processes.
15	Keyword in alt attributes	The images have alternative text that includes the keyword.	Bitmap images	Manual alt texts evaluation. <b>Rating scores</b> 0: None or only some of the images have alternative text, or the alternative text provided is not relevant to the content. 1: The images have alternative text, but it is not highly relevant. 2: The images have highly relevant alternative text.
16	Keyword in longdesc attributes	The images have a long description that includes the keyword.	Bitmap images	Manual content evaluation <b>Rating scores</b> 0: None or only some of the images have longdesc text, or the longdesc text provided is not relevant to the content. 1: The images have longdesc text, but it is not highly relevant. 2: The images have highly relevant longdesc text.
17	Keyword in <figcaption> tag	The images have a <figcaption> tag that includes the keyword.	All	Manual source code evaluation <b>Rating scores</b> 0: None or only some of the images have a <figcaption>. 1: The images have a <figcaption>, but it is not highly relevant. 2: The images have highly relevant <figcaption>.
18	Text near the images	The visualizations have nearby contextual text (immediately before or after) that includes the keyword.	All	Manual content evaluation. <b>Rating scores</b> 0: The text surrounding the images is not relevant to them. 1: The text surrounding the images is partially relevant to them. 2: The text surrounding the images is fully relevant to them.
19	IPTC metadata	The images contain IPTC properties that Google can extract.	Bitmap images	Review using an IPTC metadata viewer. <b>Rating scores</b> 0: The images do not contain IPTC metadata. 1: The images have some of the recommended IPTC metadata. 2: The images have all the recommended IPTC metadata.

ID	Factor	Description	Scope of applicability	Analytical method / Rating scores
20	max-image-preview:large tag	The page includes a max-image-preview:large tag to specify which image search engines can display a preview of the content.	Bitmap images	Manual content evaluation. <b>Rating scores</b> 0: The page does not use the tag max-image-preview:large. 1: The page uses the tag max-image-preview:large.
21	Keyword in file name	The keyword is included in the file name.	Bitmap images	Manual source code evaluation. <b>Rating scores</b> 0: The images do not have a descriptive filename. 1: The images have partially descriptive filenames. 2: The images have a fully descriptive filename.
22	Keyword in <title>, <desc> or <text> tag of vector images	Vector images include a <title> tag specifying the title of each graphic object.	Vector images	Manual source code evaluation. <b>Rating scores</b> 0: Vector images do not have a <title> tag, or it is not relevant. 1: The <title> tag of vector images is partially relevant. 2: The <title> tag of vector images is fully relevant.
23	High-quality images	Bitmap images are of sufficient quality to be displayed correctly.	Bitmap images	Manual inspection of the image. <b>Rating scores</b> 0: The quality of the image is so poor that it is impossible or very difficult to discern its content (pixelated or blurry). 1: The image is pixelated or blurred, although its content is clear. 2: The page uses high-quality images.
24	Largest Contentful Paint (LCP)	To provide a good user experience, pages should have a LCP score of 2.5 s. or less.	All	Page Speed Insights evaluation. <b>Rating scores</b> 0: > 4 seconds. 1: Between 2.5 and 4 seconds. 2: < 2.5 seconds.
25	Interaction to Next Paint (INP)	To provide a good user experience, pages should have a INP score of 200 ms. or less.	All	Page Speed Insights evaluation. <b>Rating scores</b> 0: > 500 ms. 1: between 200 ms and 500 ms. 2: < 200 ms.
26	Cumulative Layout Shift (CLS)	To provide a good user experience, pages should have a CLS score of 0.1 or less.	All	Page Speed Insights evaluation. <b>Rating scores</b> 0: > 0,25. 1: between 0,1 and 0,25. 2: < 0,1.
27	Mobile optimization	The page and its visual assets are optimized for mobile devices.	All	Qualitative manual review. <b>Rating scores</b> 0: The page is not mobile-optimized. 1: The page is partially mobile-optimized. 2: The page is fully optimized for mobile devices.

ID	Factor	Description	Scope of applicability	Analytical method / Rating scores
28	Internal backlinks	Number of internal links to the page.	All	Semrush (or similar tool) metric assessment <b>Rating scores</b> 0: Without internal inbound links. 1: Up to 5 internal inbound links. 2: > 5 internal inbound links.
29	Backlinks	Number of in-bound links to the page.	All	Semrush (or similar tool) metric assessment <b>Rating scores</b> 0: < 10000. 1: Between 10001 and 100000. 2: > 100000.
30	Referring Domains	Number of referring domains that link to the page.	All	Semrush (or similar tool) metric assessment <b>Rating scores</b> 0: < 50. 1: Between 51 y 100. 2: > 100.
31	Domain authority score	Domain authority.	All	Semrush (or similar tool) metric assessment <b>Rating scores</b> 0: < 50. 1: Between 51 y 100. 2: > 100.
32	Page authority score	Page authority.	All	Semrush (or similar tool) metric assessment <b>Rating scores</b> 0: < 20. 1: Between 21 y 50. 2: > 50.
33	Monthly visits	Estimated monthly domain-level visits.	All	Semrush (or similar tool) metric assessment <b>Rating scores</b> 0: < 1000000. 1: Between 1000001 y 10000000. 2: > 10000000.
34	Domain Organic Traffic	Estimated monthly domain-level traffic.	All	Semrush (or similar tool) metric assessment <b>Rating scores</b> 0: < 500000. 1: Between 500001 y 5000000. 2: > 5000000.

### 3. Conclusions

As a result of this proposal, a total of 34 factors were identified and classified as follows: content SEO (7 factors); technical SEO (7 factors); EEAT (5 factors); SEO for graphic components (9 factors); and off-page SEO (6 factors).

SEO and journalism have joined forces for years, ensuring that media outlet content attains the visibility necessary to remain relevant in the internet age. From the most powerful and well-funded media outlets to the most local and modest ones, organisations have incorporated dedicated personnel and practices related to search engine optimisation. Narrative data visualisations represent a distinct byproduct that demands particular attention, as their

optimisation requires the alignment of various factors. This is especially crucial for images, charts, infographics, and other graphic components, as well as for broader aspects influencing the journalistic field, such as EEAT. The proposed SEOvisFrame outlines a set of indicators for assessing the effectiveness of SEO in storytelling data visualisations. It can also serve as a reference guide or checklist to be used during the conceptualisation and design of the communication product.

## 4. Funding

This work is part of the Project “Parameters and strategies to increase the relevance of media and digital communication in society: curation, visualisation and visibility (CUVICOM)”. Grant PID2021-123579OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU.

## 5. References

- Alcaraz Martínez, R. (2022). SEO centrado en el usuario: un cambio de paradigma orientado a satisfacer mejor las necesidades de las personas en el buscador de Google. *Cuadernos de Documentación Multimedia*, 32. <https://doi.org/10.5209/cdmu.81156>
- Alcaraz Martínez, R. (2024). Integración de la optimización de la experiencia de búsqueda (SXO), la usabilidad, la arquitectura de la información y la accesibilidad web: explorando un terreno común para mejorar la visibilidad en buscadores. *Ibersid*. 18(1). <https://doi.org/10.54886/ibersid.v18i1.4966>
- Alexander, S., & Vetere, C. (2011). Telling the data story the right way. *Healthcare Financial Management*, 65(10), 104-110.
- Almukhtar, F., Nawzad, M., & Shahab, K. (2021). Search engine optimization: A review. *Applied Computer Science*, 17(1), 70–80. <https://doi.org/10.35784/acs2021-07>
- Chotikitpat, K., Nilsook, P., & Sodsee, S. (2015). Techniques for improving website rankings with search engine optimization (SEO). *Advanced Science Letters*, 21(10), 3219–3224. <https://doi.org/10.1166/asl.2015.6503>
- Córdoba-Cabús, A. (2020). Estándares de calidad en el periodismo de datos: fuentes, narrativas y visualizaciones en los Data Journalism Awards 2019. *Profesional de la Información*, 29(3), e290328. <https://doi.org/10.3145/epi.2020.may.28>
- Córdoba-Cabús, A., García-Borrego, M., & López-Martín, Á. (2020). El periodismo de datos durante la crisis sanitaria del Covid-19 en la prensa española. *Revista Ibérica de Sistemas e Tecnologías de Informação*, 35, 325-337. <https://hdl.handle.net/10630/20031>
- Dean, B. (2024). *Google’s 200 ranking factors: the complete list*. <https://backlinko.com/google-ranking-factors>
- Dick, M. (2011). Search engine optimisation in UK news production. *Journalism Practice*, 5(4), 462-477. <https://doi.org/10.1080/17512786.2010.551020>
- Domínguez, E. (2016). El reto narrativo de integrar datos en los formatos audiovisuales interactivos. In E. Blanco Castilla & M. Quesada (Coords.). *Periodismo de Datos* (pp. 87-105). Cuadernos Artesanos de Comunicación. <http://www.cuadernosartesanos.org/2016/cac112.pdf>

Enge, E.; Spencer, S.; Stricchiola, J. (2023). *The art of SEO: Mastering search engine optimization*. O'Reilly.

Escandell-Poveda, R., Iglesias-García, M., & Papí-Gálvez, N. (2021). Who does SEO in Spain? a cybermetric methodology for the construction of company universes. *Profesional de la Información*, 30(4). <https://doi.org/10.3145/epi.2021.jul.19>

Ferraz, R. (2017). Accessibility and search engine optimization on scalable vector graphics. In *2017 IEEE 4th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, (pp. 94-98). <https://doi.org/10.1109/ISCMI.2017.8279605>

Ferreras Rodríguez, E. M. (2016). El periodismo de datos en España. *Estudios sobre el Mensaje Periodístico*, 22(1) 255-272. [https://doi.org/10.5209/rev\\_ESMP.2016.v22.n1.52594](https://doi.org/10.5209/rev_ESMP.2016.v22.n1.52594)

García-Carretero, L., Codina, L., Díaz-Noci, J., & Iglesias-García, M. (2016). Herramientas e indicadores SEO: características y aplicación para análisis de cibermedios. *Profesional de la Información*, 25(3), 497–504. <https://doi.org/10.3145/epi.2016.may.19>

Giomelakis, D. (2023). Semantic search engine optimization in the news media industry: Challenges and impact on media outlets and journalism practice in Greece. *Social media+ society*, 9(3), 20563051231195545.

Giomelakis, D., & Veglis, A. (2015). Employing search engine optimization techniques in online news articles. *Studies in Media and Communication*, 3(1), 22-33. <https://doi.org/10.11114/smc.v3i1.683>

Google (2014). *HTTPS as ranking signal*. Google search central. <https://developers.google.com/search/blog/2014/08/https-as-ranking-signal>

Google (2024a). *Google search essentials*. <https://developers.google.com/search/docs/essentials>

Google (2024b). *General guidelines*. <https://static.googleusercontent.com/media/guidelines.raterhub.com/es//searchqualityevaluatorguidelines.pdf>

Google (2024c). *Google image SEO best practices*. <https://developers.google.com/search/docs/appearance/google-images>

Google (2025a). *General structured data guidelines*. <https://developers.google.com/search/docs/appearance/structured-data/sd-policies>

Google (2025b). *Image metadata in Google images*. <https://developers.google.com/search/docs/appearance/structured-data/image-license-metadata>

Krstić, N. (2019). The relationship between website management and search engine optimisation. *Megatrend Revija*, 16(2), 155-168. <https://scindeks-clanci.ceon.rs/data/pdf/1820-3159/2019/1820-31591902155K.pdf>

Lewandowski, D. (2023). Search Engine Optimization (SEO). In *Understanding search engines* (pp. 175-190). Springer. [https://doi.org/10.1007/978-3-031-22789-9\\_9](https://doi.org/10.1007/978-3-031-22789-9_9)

Lopezosa, C., Codina, L., & Caldera Serrano, J. (2018). SEO semántico: Framework ISS para la optimización de sitios intensivos en contenidos. *Cuadernos de Documentación Multimedia*, 29, 97-123. <https://doi.org/10.5209/CDMU.60607>

- Lopezosa, C., Codina, L., & Gonzalo-Penela, C. (2019). SEO off page y construcción de enlaces: estrategias generales y transmisión de autoridad en cibermedios. *Profesional de la Información*, 28(1). <https://doi.org/10.3145/epi.2019.ene.07>
- Lopezosa, C., Codina, L., Gonzalo Penela, C., & Pont Sorribes, C. (2020). SEO y periodismo: Marco de optimización global como parte del emprendimiento en cibermedios. <http://hdl.handle.net/10230/44923>
- Maciá Domene, F. (2019). *SEO avanzado: casi todo lo que sé sobre posicionamiento web*. Anaya Multimedia
- Marcos, M. C., Gavin, F., & Arapakis, I. (2015, September). Effect of snippets on user experience in web search. In *Proceedings of the XVI international conference on human computer interaction* (pp. 1-8). <https://doi.org/10.1145/2829875.2829916>
- Mueller, J. (2015). *English Google webmaster central office-hours hangout*. [https://youtu.be/i\\_xnKznRNCc?t=24m5s](https://youtu.be/i_xnKznRNCc?t=24m5s)
- Peiró, K., & Guallar, J. (2013). Introducció al periodisme de dades. Característiques, panoràmica i exemples, *Ítem: Revista de Biblioteconomia i Documentació*, 57, 23-37. <https://raco.cat/index.php/Item/article/view/269701>
- Radcliffe, D., & Lewis, S. C. (2019). The datafication of journalism: strategies for datadriven storytelling and industry-academy collaboration. In *The data journalism handbook 2*. <https://datajournalism.com/read/handbook/two/training-datajournalists/the-datafication-of-journalism>
- Rodas, C. M., Vidotti, S. A. B. G., & Monteiro, S. D. (2016). Interface de busca do Google e Yahoo: a experiência do usuário sob o olhar do eye tracking. *Informação & Sociedade*, 26(2). <https://doi.org/10.22478/ufpb.1981-0695.2017v12n1.34043>
- Salem, H., Salloum, H., Orabi, O., Sabbagh, K., & Mazzara, M. (2025). Enhancing News Articles: Automatic SEO Linked Data Injection for Semantic Web Integration. *Applied Sciences*, 15(3), 1262. <https://doi.org/10.3390/app15031262>
- Seyser, D., & Zeiller, M. (2018). Scrollytelling: An analysis of visual storytelling in online journalism. In *22nd International Conference Information Visualisation* (pp. 401-406). <https://doi.org/10.1109/IV.2018.00075>
- Semrush (2024). *Google ranking factors*. <https://go.semrush.com/Ranking-Factors.html>
- Vicente, J.E., Saquete, R., Rodríguez, R., & Maciá Domene, F. (2024). *SEO técnico: aprende a optimizar tu web como un profesional*. Anaya Multimedia.
- Ziakis, C., Vlachopoulou, M., Kyrkoudis, T., Karagkiozidou, M. (2019). Important factors for improving Google search rank. *Future Internet*, 11(32). <https://doi.org/10.3390/fi11020032>



# **Semantic content and information visualization: A proposal for a data-driven communication narrative**

**Mario Pérez-Montoro**

Universitat de Barcelona, Spain

<https://orcid.org/0000-0003-2426-8119>

---

Pérez-Montoro, M. (2025). Semantic content and information visualization: A proposal for a data-driven communication narrative. In J. Guallar, M. Váñez, & A. Ventura-Cisquella (Coords). *Digital communication. Trends and good practices* (pp. 85-103). Ediciones Profesionales de la Información. <https://doi.org/10.3145/cuvicecom.07.eng>

---

## Abstract

Charts and graphs are one of the main visual tools used to represent and convey information. A significant part of our daily work involves searching for, obtaining, and refining a dataset that allows us to tell a story. However, data does not speak for itself. It needs to be visualized to reveal the underlying analysis and bring to light the patterns it contains. In this work, we aim to characterize the main communicative contexts and content types that can be conveyed through graphs. We will also explore the most suitable visual approaches to efficiently and effectively fulfill these communicative purposes. In other words, in this chapter, we seek to present a structured methodology commonly used to properly implement effective information visualization. First, we identify the type of information to be encoded; second, we choose the type of graph that best represents that information; and finally, we design the visual elements that compose the graph.

## Keywords

Information visualization; Data visualization; Visual information communication; Information design; Graphs; Charts; Visual representation of information.

---

## 1. Introduction

Charts and graphs are one of the main visual tools used to represent and communicate information. A significant part of our daily work involves searching for, obtaining, and refining datasets that allow us to tell a story. However, data does not speak for itself; it needs to be visualized to reveal derived analyses and uncover the patterns it holds.

In this work, we aim to characterize the variety of contexts and communicative purposes that can be conveyed using a graph or chart. We will also explore which visual approach is most suitable to efficiently and effectively fulfill these communicative intentions. In other words, this chapter seeks to systematically present the methodology commonly used to implement effective information visualization. First, the identification of the type of information to be encoded; second, the selection of the type of graph or chart that best represents that information; and finally, the design of the visual elements of the graph.

To achieve this goal and following this structure, we will address the following points. In the first section (Section 2), we will discuss the main basic types of information or semantic content that can be represented in a graph or chart, illustrating them with examples organized into tables for easier understanding. In Section 3, we will examine which type of visualization is best suited to communicate each of these types of informational content.

## 2. Semantic content

Intuitively, any communication strategy is based on three basic elements: the message or semantic content to be conveyed, the intended audience, and the context surrounding the communicative act.

Communication through graphs or charts is no different. It depends on the message or informational content we aim to deliver to our audience, who we want to reach, and the context

(the document where the graph is embedded, accompanying text, transmission channel, communicative intent, or the timing of its consumption, among many other factors) that frames the communication effort. For now, let us focus exclusively on the content we wish to convey.

Typically, the semantic content communicated through a graph is a proposition (an idea, in a non-technical sense) asserting the existence of a relationship between two or more types of information. Specifically, it asserts a relationship between two (or more) items of quantitative and/or qualitative information. In other words, it expresses some kind of relationship between values belonging to two or more quantitative variables, between the categories or attributes of two or more qualitative variables, or between categories or attributes and quantitative values.

In this sense, it is possible to classify the main types of relationships represented in the semantic content of graphs (Zelazny, 2001; Few, 2012; Shneiderman, 1996; Pérez-Montoro, 2022). These include, for example, nominal comparison, time series, ranking, part-to-whole, deviation, distribution, and correlation relationships.

## 2.1. Nominal comparison

Nominal comparison relationships could be considered the most basic and common semantic content represented using graphs.

From a technical perspective, this semantic content is understood as a proposition (an idea) that establishes a relationship in which a quantitative value is assigned or connected to each of the items (values or categories) of a qualitative variable to facilitate the comparison of those numeric values associated with those items.

**Figure 1**  
*Job positions and salaries of ACME company employees.*

Job position	Salary
Manager	150.000
Analyst	40.000
Documentalist	40.000
Administrative Assistant	32.000
Junior Staff	25.000
Intern	8.000

(Own elaboration, simulated data)

An example of this type of semantic content can be seen in Figure 1. This figure presents a table illustrating the relationship between the various items, categories, or values (manager, analyst, documentalist, administrative assistant, junior staff, and intern) of a qualitative variable (job positions or profiles in ACME company) and a quantitative value (gross salary received by employees based on their roles).

## 2.2. Time series

Time series relationships are another common type of semantic content often communicated using graphs.

Conceptually, this content is identified as a proposition (an idea) that establishes a relationship where a quantitative value is assigned or connected to each of the items (values or categories) of a temporal variable, facilitating the comparison of those numeric values associated with those time units.

**Figure 2**  
*Quarterly sales during the previous fiscal year.*

Quarter	Sales
First	456.876
Second	189.387
Third	63.829
Fourth	122.345

(Own elaboration, simulated data)

An example of this type of semantic content can be found in Figure 2. The figure presents a table showing the relationship between the different items, categories, or values (first, second, third, and fourth) of a temporal variable (quarters of the previous year) and a quantitative value (sales generated by the company during those time periods).

### 2.3. Ranking

This semantic content refers to a proposition (an idea) that establishes a relationship in which a quantitative value is assigned or connected to each item (values or categories) of a qualitative variable, and this assignment of numeric values creates an order (from highest to lowest or vice versa) among the items of the qualitative variable.

**Figure 3**  
*Sales by province in Catalonia.*

Ranking	Province	Sales
1	Barcelona	456.876
2	Tarragona	189.387
3	Gerona	122.345
4	Lérida	63.829

(Own elaboration, simulated data)

Figure 3 illustrates this type of semantic content. It shows a table that represents the relationship between the various items, categories, or values (Barcelona, Tarragona, Girona, and Lleida) of a qualitative variable (the provinces of Catalonia) and a quantitative value (sales achieved in each province). This assignment of numeric values establishes an order among the provinces.

### 2.4. Part-to-whole

The part-to-whole relationship is another common type of semantic content often represented using graphs. Conceptually, this semantic content refers to a proposition (an idea) that establishes a relationship where a quantitative value is assigned or connected to each item (values or categories) of a qualitative variable. It highlights that the aggregation of all these

qualitative items forms a whole or entity. Thus, each of these categories is identified as a part of that whole. It is also worth noting that percentages are the most commonly used unit of measurement for expressing the contribution of parts to a whole, where the whole is represented as 100%, with each part as a portion of that total.

**Figure 4**

*Activities carried out during a day.*

Activity	Percentage
Eating	7,40%
TV/Internet	8,20%
Socializing	9,40%
Sleeping	29,20%
Working	45,80%
Total	100,00%

(Own elaboration, simulated data)

An example of this type of semantic content is shown in Figure 4, which is a table illustrating the relationship between various items, categories, or values (working, eating, socializing, watching TV/Internet, and sleeping) of a qualitative variable (different activities performed throughout the day) and a quantitative value (the percentage of daily time allocated to each activity). This numerical assignment expresses the quantitative relationship of each part (its weight, so to speak) relative to the whole and facilitates comparisons between these numerical values.

## 2.5. Deviation

Intuitively, the deviation relationship expresses how a set of quantitative data varies relative to a specific numeric reference value.

From a technical perspective, this semantic content is a proposition (an idea) that establishes a relationship where a quantitative value is assigned or connected to each item (values or categories) of a qualitative variable. These quantitative values are compared to a specific numeric reference value, which is typically assigned to a qualitative item (from the same or another qualitative variable). In some cases, there may be two or more reference values for comparison.

**Figure 5**

*Comparison of last year's sales between sector competitors and our company.*

Company	Sales
Competitor A	280%
Competitor B	220%
Competitor C	115%
Competitor D	60%
Competitor E	20%

(Own elaboration, simulated data)

An example of this type of semantic content is presented in Figure 5, which is a table illustrating the relationship between various items, categories, or values (Competitor A, Competitor B, Competitor C, Competitor D, and Competitor E) of a qualitative variable (different companies competing in the same sector as ours) and a quantitative value (the percentage of sales of each company relative to our company's sales). This numeric assignment expresses the quantitative relationship between each company's sales and ours, facilitating comparisons between these values. In this case, the reference value (our company's sales) is identified as 100%, with the remaining values (competitors' sales) expressed as proportions or percentages relative to this primary reference value.

## 2.6. Distribution

The distribution relationship is a type of semantic content corresponding to a proposition (an idea) that establishes a relationship where a quantitative value is assigned or connected to each item (values or categories) of an interval variable. This facilitates comparisons between these numeric values associated with those interval units. This content expresses how the quantitative values of a dataset are distributed or spread over a range (the interval between the minimum and maximum values), from the lowest to the highest, across the categories of that interval variable.

**Figure 6**

*Number of patients by age group in a population of epilepsy patients.*

Age group	Patients
1-10	15
11-20	30
21-30	55
31-40	75
41-50	25

(Own elaboration, simulated data)

An example of this type of semantic content is shown in Figure 6. The figure presents a table illustrating the relationship between various items, categories, or values (0–15, 16–30, 31–45, 46–60, and 61–70) of an interval variable (age ranges in a population of epilepsy patients) and a quantitative value (the number of patients). This numeric assignment expresses the distribution of the number of patients (how they are spread) across age ranges in a population of 200 epilepsy patients.

## 2.7. Correlation

Intuitively, the correlation relationship seeks to express whether two sets of quantitative values are related and, if so, how one changes relative to the other.

From a more conceptual perspective, and in its simplest form, this semantic content is a proposition (an idea) that establishes a relationship where two quantitative values (each belonging to a different quantitative variable) are assigned or connected to each item (values or categories) of a qualitative variable. This facilitates the identification of patterns in the variation between these two types of numeric values. It seeks to determine whether a systematic relationship exists between the values of one quantitative variable and those of another.



**Figure 7**

*Investment and profits by manufactured product.*

Product	Investment	Profits
Product 01	22.864	678.543
Product 02	45.789	778.765
Product 03	50.678	783.213
Total	119.331	2.240.521

(Own elaboration, simulated data)

An example of this type of semantic content is presented in Figure 7. The figure shows a table illustrating the relationship between various items, categories, or values (Product 1, Product 2, and Product 3) of a qualitative variable (different products manufactured by ACME company) and two numeric values, each belonging to a different quantitative variable (investment in the production and distribution of a product, and total profits from its commercial life).

### 3. Visualization of content

Once we have reviewed the main semantic content we aim to communicate through graphics, the next step is to address the process of visualizing these elements or propositional units using graphical representation. In other words, we need to determine the best graphical solutions to effectively encode each type of semantic content.

One critical aspect of creating an effective visualization is selecting the chart that best communicates the information efficiently and effectively. This is not a straightforward task. It depends on several factors, particularly the type of information being represented and the communicative intent. Each type of semantic content requires a specific graphical representation, depending on the communication objectives.

Most data management programs do not provide guidance on selecting the most appropriate chart for a given visualization (Hugues and van Dam, 2013; Hearn, 2011; Shirley, 2009). Only a few software tools offer limited (and often incomplete) recommendations based on the structure of the data being represented (Cherven, 2015; Jones, 2014; Khan, 2016).

In this section, we aim to address this issue. We will provide a set of resources and guidelines to help you select the most communicatively appropriate chart for representing different types of semantic content. Specifically, we will introduce the best practices for visualizing nominal comparisons, time series relationships, rankings, part-to-whole relationships, deviations, distributions, and correlations.

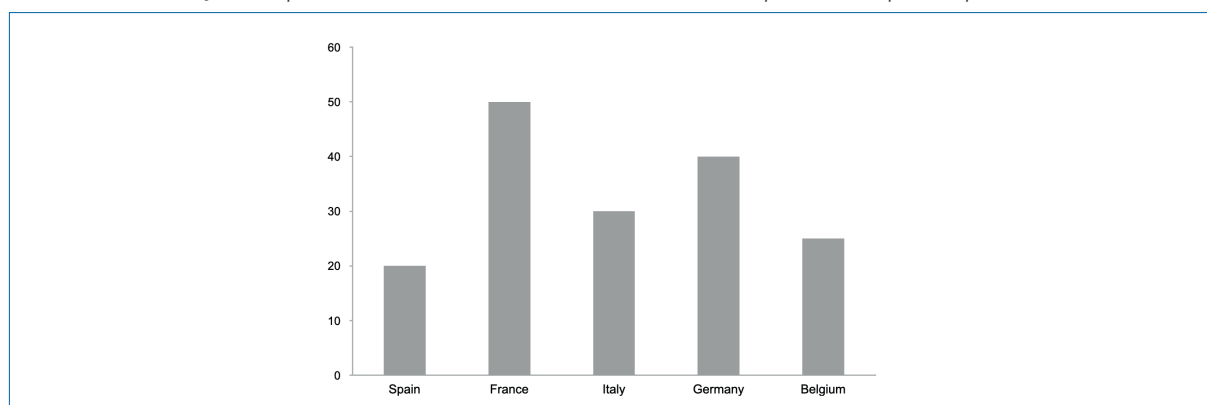
#### 3.1. Visualizing nominal comparison

A chart that accurately represents this type of semantic content must use visual elements that simultaneously encode numeric values and their associated qualitative items. It must also ensure that each resulting pair (quantitative value and qualitative category) is independent, with no intrinsic connection between the pairs.

Given this semantic requirement, the best visual elements for this relationship are bars (or columns) (Figure 8) and points. These elements display the pairs (quantitative value and qualitative item) as visually independent, thus facilitating comparisons between the numeric values.

**Figure 8**

*Goals conceded by the top five national football teams in the last European Championship.*

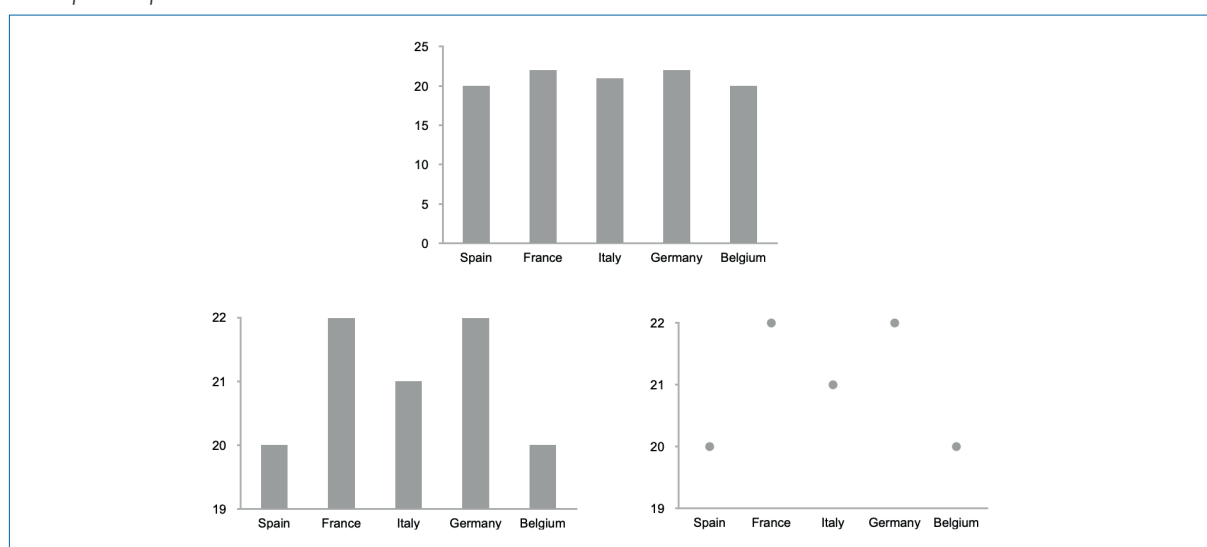


(Own elaboration, simulated data)

In certain contexts, point charts can be a good alternative to column (or bar) charts for this semantic content—specifically when the differences between quantitative values are minimal or when the range of values is narrow, indicating low variability.

**Figure 9**

*Column and point charts showing goals conceded by the top five national football teams in the last European Championship.*



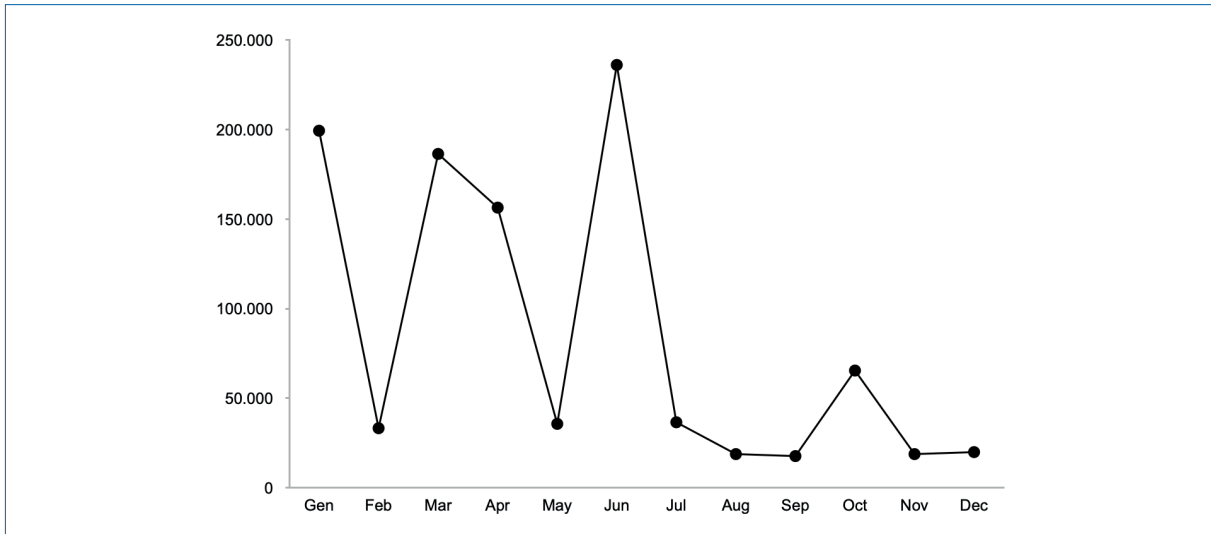
(Own elaboration, simulated data)

Using a column (or bar) chart in these cases might make it harder to perceive differences in the lengths of the bars, thus complicating visual comparisons. Attempting to address this by starting the vertical (or horizontal) axis at a value greater than zero is an incorrect practice. Bars encode quantitative values based on their length, and this requires starting the axis at zero. Point charts, by contrast, avoid this issue because quantitative and qualitative values are encoded by their position on the X and Y axes (Figure 9).

### 3.2. Visualizing time series

Time series relationships link a quantitative value to each item (value or category) of a temporal variable, facilitating the comparison of numeric values associated with these temporal units.

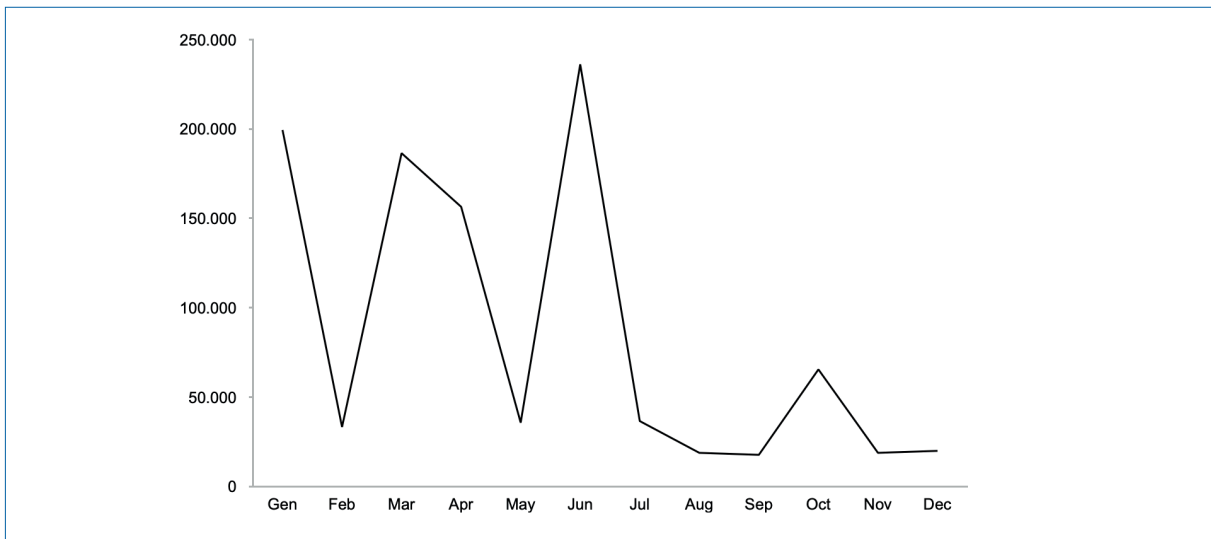
**Figure 10**  
*Monthly sales of ACME during the previous year.*



(Own elaboration, simulated data)

A chart for this type of semantic content must encode numeric values alongside their associated temporal categories. One axis should represent the temporal variable, with labels for each time unit (year, month, etc.), while respecting the natural order of time. In Western cultures, this order is conventionally displayed horizontally, from left to right, along the X-axis.

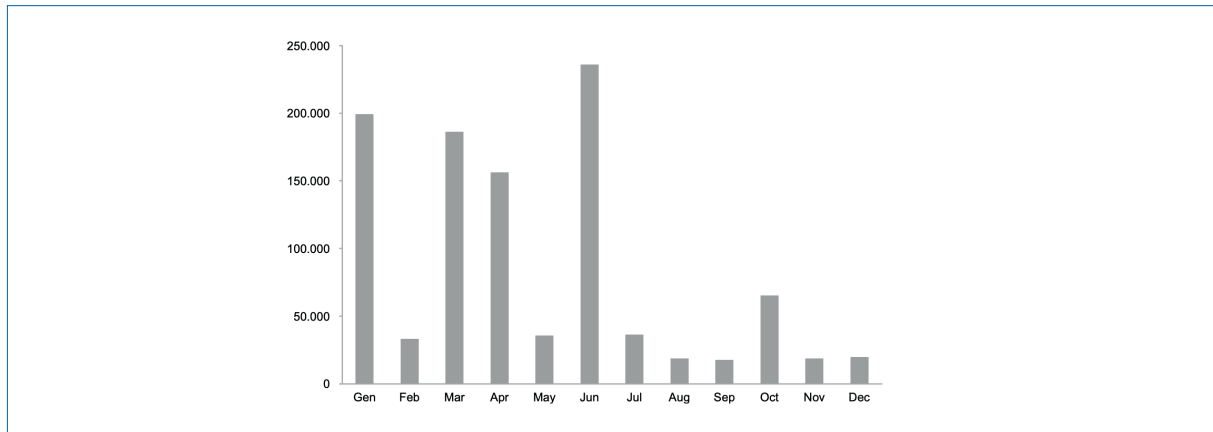
**Figure 11**  
*Monthly sales of ACME during the previous year.*



(Own elaboration, simulated data)

The best visual elements for this relationship are combinations of points and lines (Figure 10), standalone lines (Figure 11), and columns. These elements connect each pair (quantitative value and temporal item) to the next, illustrating the temporal continuity.

**Figure 12**  
*Monthly sales of ACME during the previous year.*



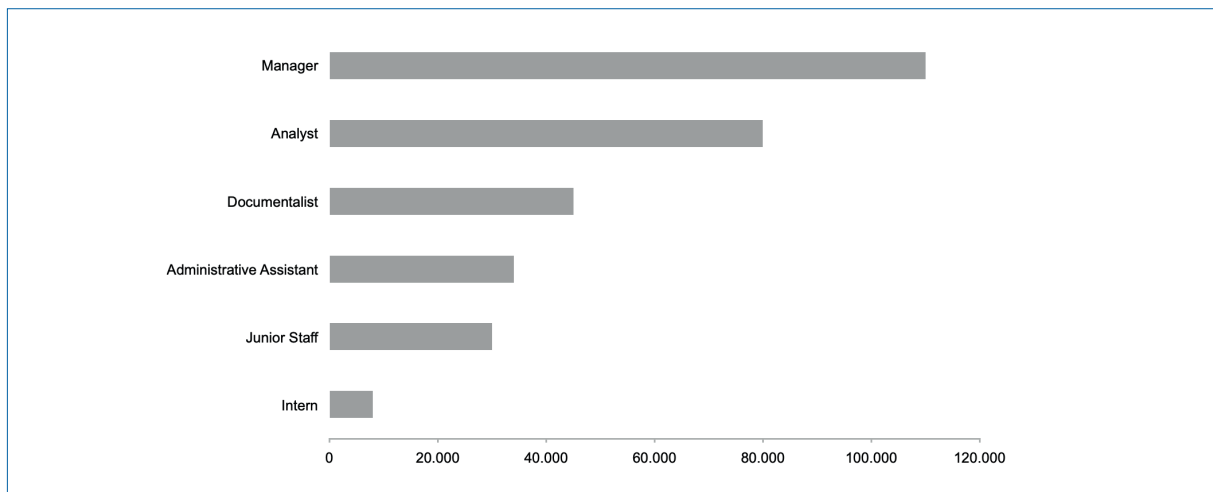
(Own elaboration, simulated data)

Column charts are better suited when the focus is on individual quantitative values associated with specific time points, rather than on the overall trend of the series (Figure 12).

### 3.3. Visualizing ranking

Ranking relationships link a quantitative value to each item (value or category) of a qualitative variable, establishing an order (ascending or descending) among the items.

**Figure 13**  
*Salary by job position at ACME*

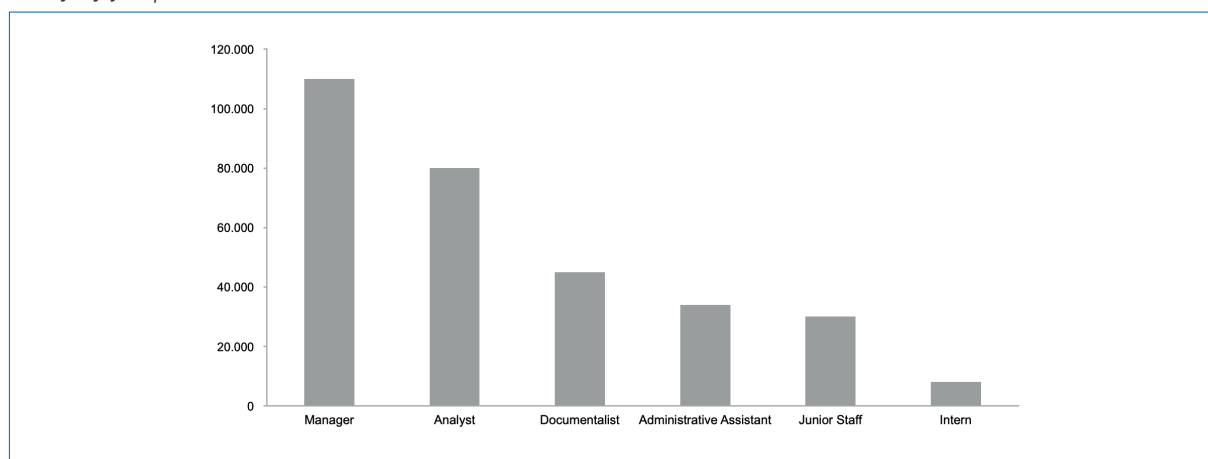


(Own elaboration, simulated data)

A chart that accurately represents this type of semantic content must encode numeric values and their associated categories while visually displaying the order of the qualitative items. This requires placing the qualitative variable along one axis in the correct order, with the quantitative scale on the other.

**Figure 14**

*Salary by job position at ACME.*

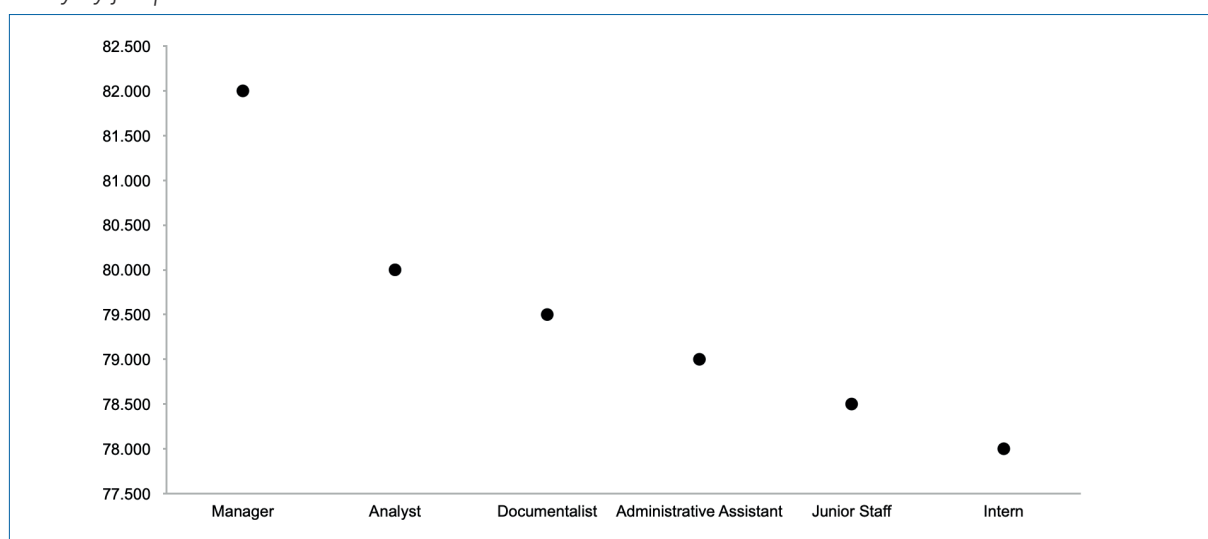


(Own elaboration, simulated data)

Bars (Figure 13), columns (Figure 14), and points are the most suitable visual elements for this type of relationship. They encode pairs (quantitative value and qualitative item) while visually displaying the order of the qualitative items.

**Figure 15**

*Salary by job position at ACME.*



(Own elaboration, simulated data)

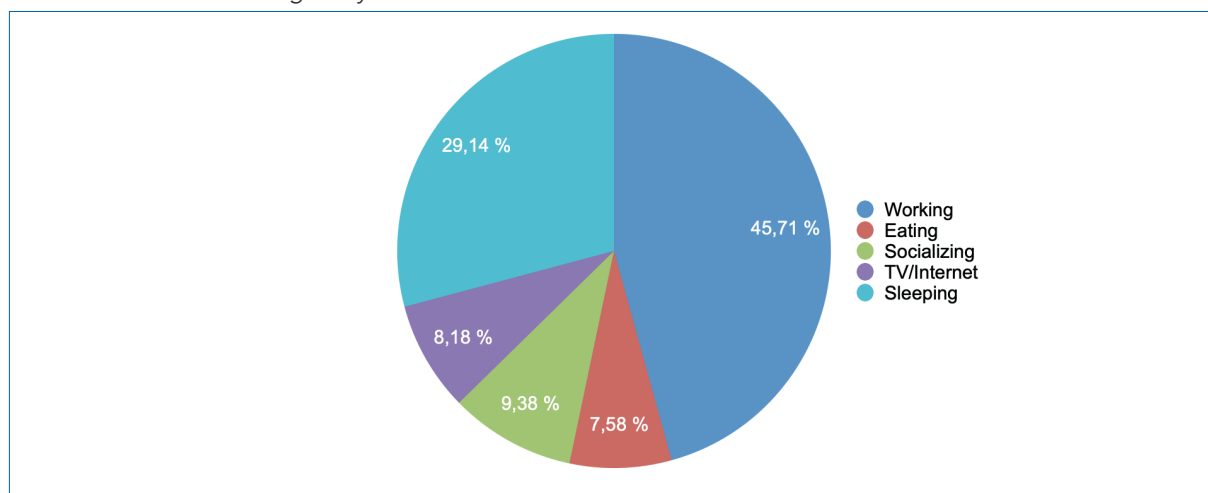
Point charts are preferable in cases where the differences between quantitative values are minimal, making it easier to visualize rankings clearly (Figure 15).

### 3.4. Visualizing part-to-whole

Part-to-whole relationships link quantitative values to qualitative items that collectively form a whole. A suitable chart must encode these values and categories while illustrating the quantitative relationship between the parts and the whole, facilitating comparisons.

In many communication contexts, pie charts and stacked bar (or column) charts are commonly used.

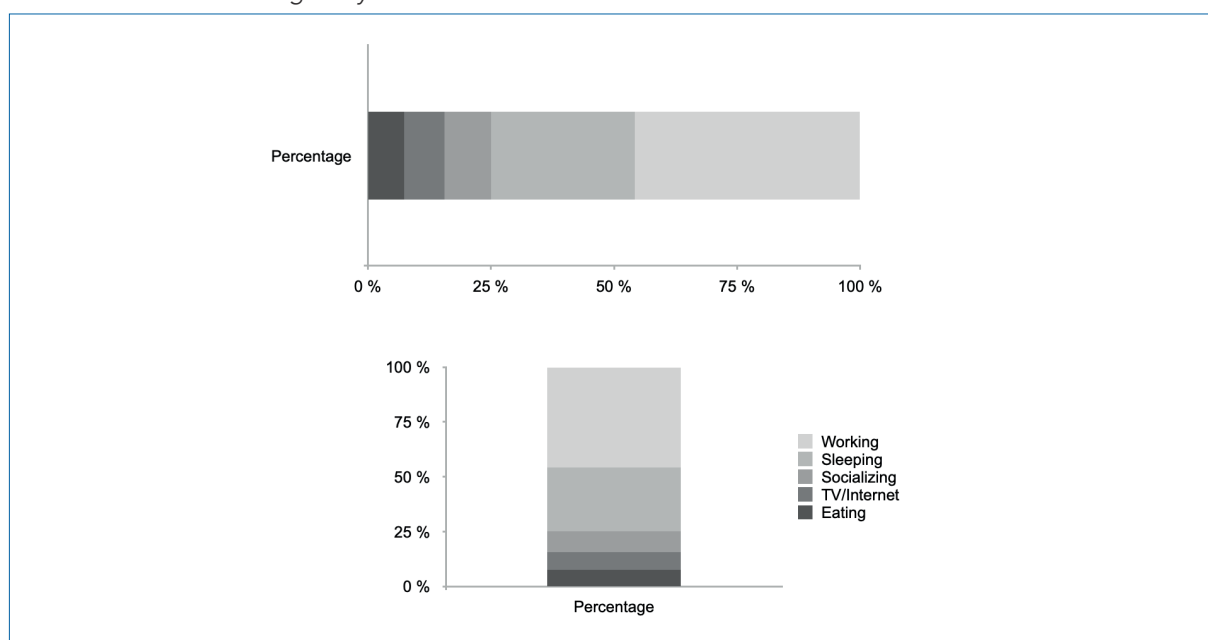
**Figure 16**  
 Activities carried out during a day.



(Own elaboration, simulated data)

In pie charts, each segment represents a quantitative value (based on the angle between radii) associated with a qualitative item (indicated by color). These charts highlight the relationships between parts and the whole (Figure 16).

**Figure 17**  
 Activities carried out during a day.



(Own elaboration, simulated data)

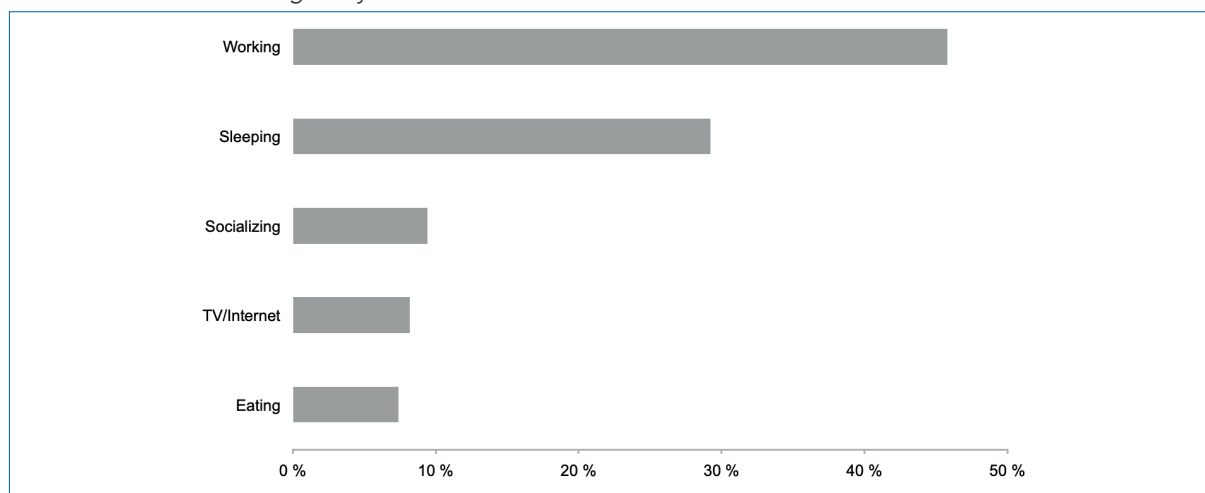
Stacked bar (or column) charts represent the whole as a rectangular bar, with individual segments showing the parts. These charts similarly convey the relationships between parts and the whole (Figure 17).

However, unstacked bars (Figure 18) and columns (Figure 19) are often the most effective visual elements for representing part-to-whole relationships, as they avoid the visual limitations of pie and stacked charts.



**Figure 17**

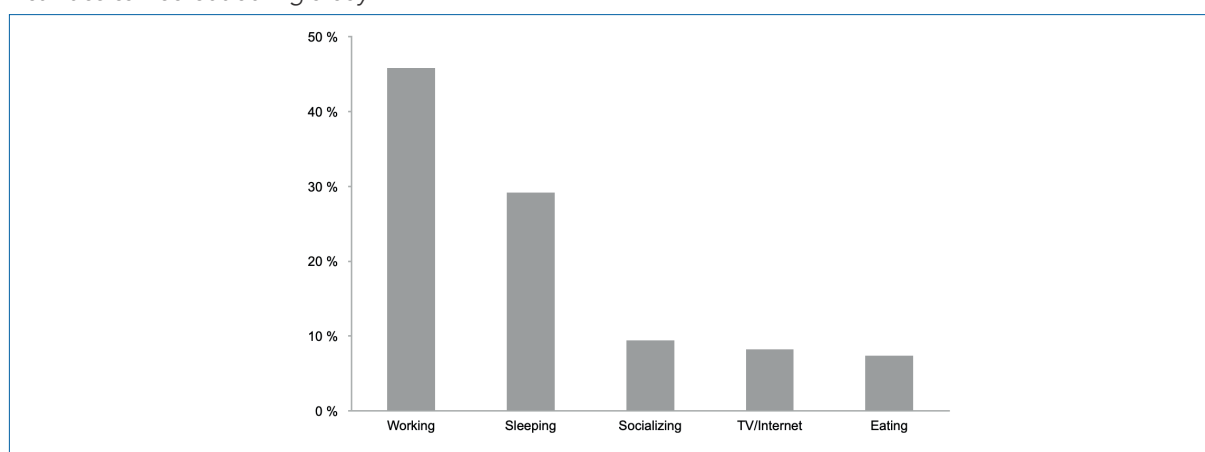
*Activities carried out during a day.*



(Own elaboration, simulated data)

**Figure 19**

*Activities carried out during a day.*



(Own elaboration, simulated data)

It is worth noting that lines and points (or combinations of lines and points) are not suitable for visualizing part-to-whole relationships.

### 3.5. Visualizing deviation

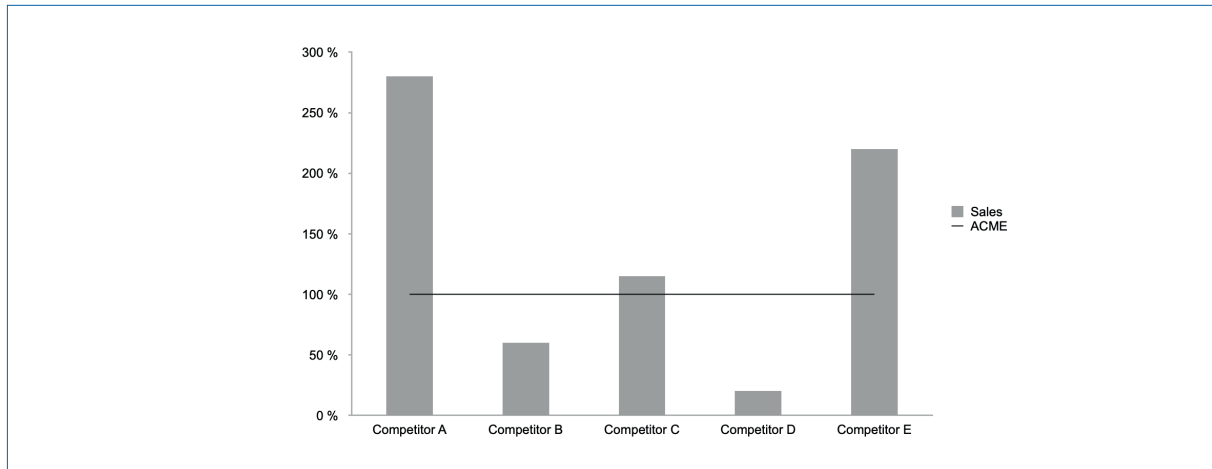
As previously stated, a deviation relationship assigns or connects a quantitative value to each item (value or category) of a qualitative variable. This relationship compares the associated quantitative values to a specific numeric reference value.

A chart that effectively represents this type of semantic content must use visual objects that simultaneously encode the numeric value and the associated category, while enabling quantitative comparison of those numeric data points against the reference value.

The visual objects that best encode this relationship are bars and columns as they adequately represent these pairs (quantitative value and qualitative item) and visually illustrate the quantitative relationship between the values and the reference value.

**Figure 20**

*Comparison of last year's sales between sector competitors and our company.*

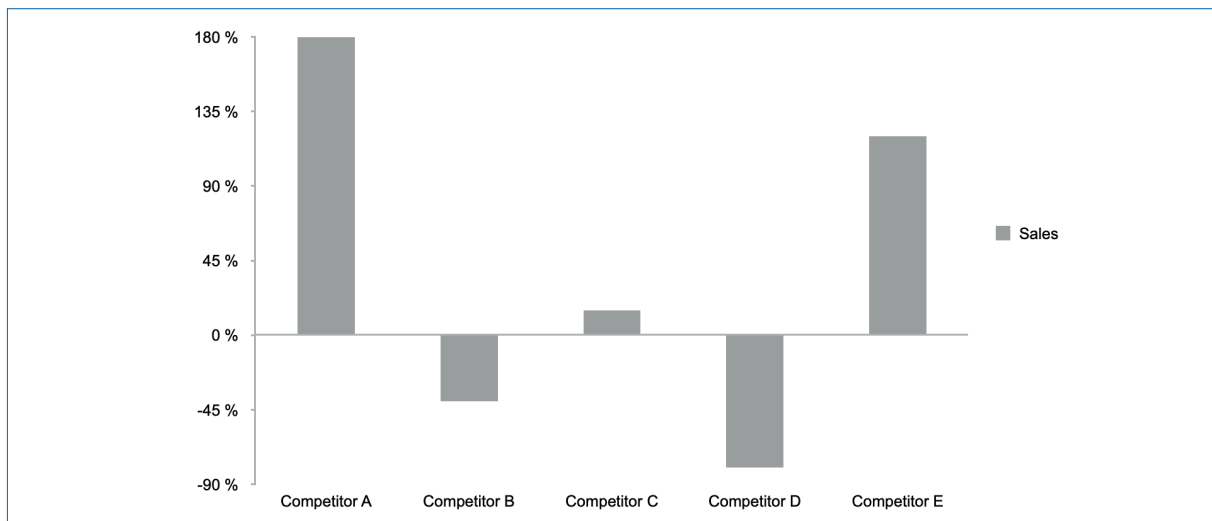


(Own elaboration, simulated data)

In a bar chart (or column chart), each rectangular figure encodes a quantitative value (along the X-axis for bars and the Y-axis for columns) associated with a qualitative item (along the Y-axis for bars and the X-axis for columns). If the reference value is represented with a line or aligned with one of the axes (the Y-axis for bars and the X-axis for columns), it also conveys the quantitative relationship between the numeric values and the reference value (Figure 20).

**Figure 21**

*Comparison of last year's sales between sector competitors and our company.*



(Own elaboration, simulated data)

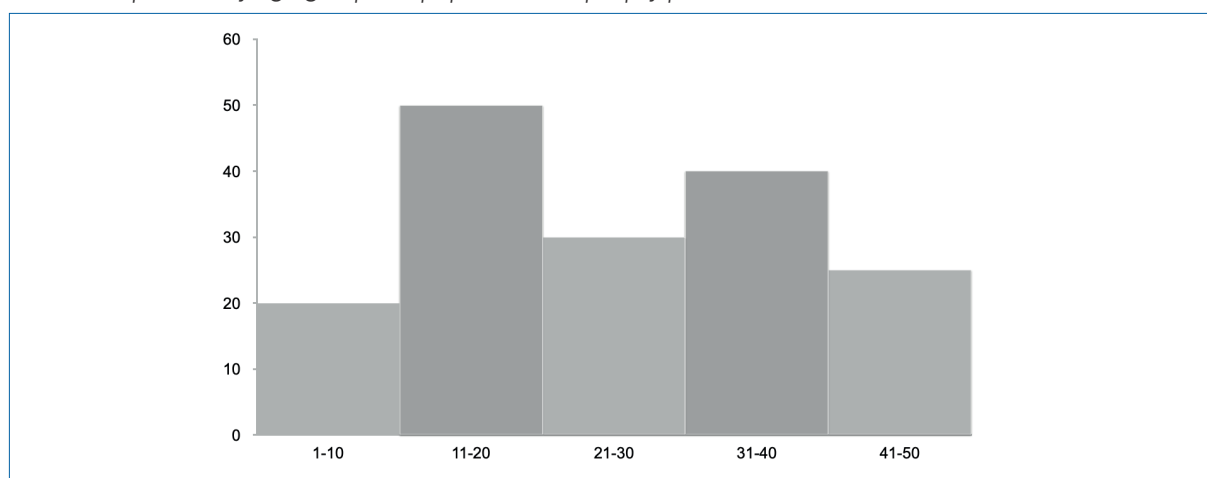
As an alternative to the mixed graph, we can provide a version without a line, where the columns are the only visual objects. In this case, to avoid using a line, the reference value (ACME's sales) is aligned with the X-axis (Figure 21).

### 3.6. Visualizing distribution

A distribution relationship, as previously defined, assigns or connects a quantitative value to each item (value or category) of an interval variable, facilitating the comparison of numeric values across the interval units.

**Figure 22**

*Number of patients by age group in a population of epilepsy patients .*

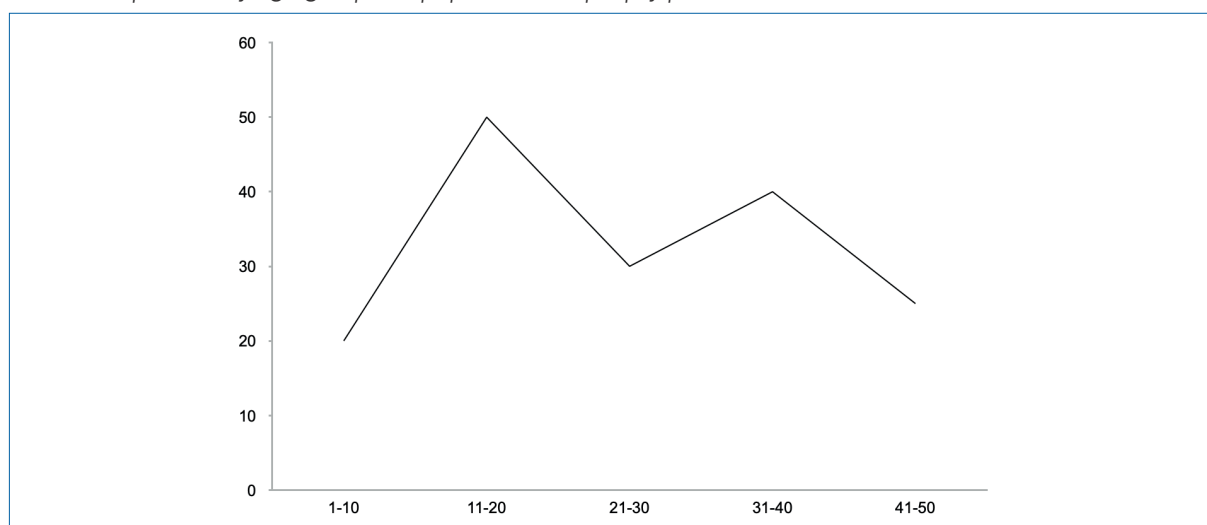


(Own elaboration, simulated data)

A chart that effectively represents this type of semantic content must use visual objects that simultaneously encode the numeric value and the associated category while illustrating how the numeric values are distributed across the range of the set they belong to, through the categories that comprise the interval variable.

**Figure 23**

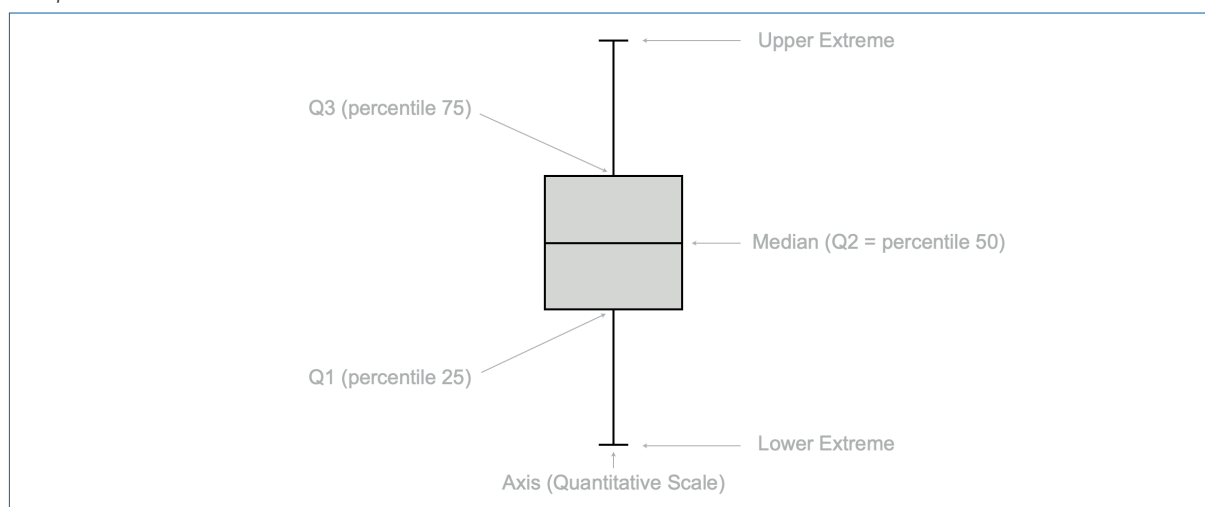
*Number of patients by age group in a population of epilepsy patients.*



(Own elaboration, simulated data)

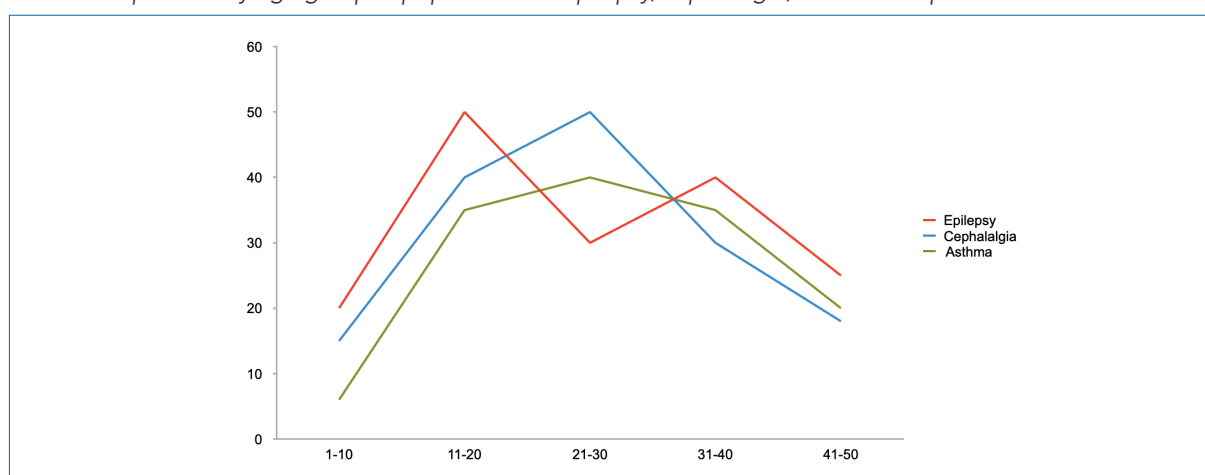
An additional consideration is whether we want to represent the distribution of a single data-set or visualize the distributions of two or more datasets within the same chart.

**Figure 24**  
Box plot.



(Own elaboration, simulated data)

**Figure 25**  
Number of patients by age group in populations of epilepsy, cephalalgia, and asthma patients.

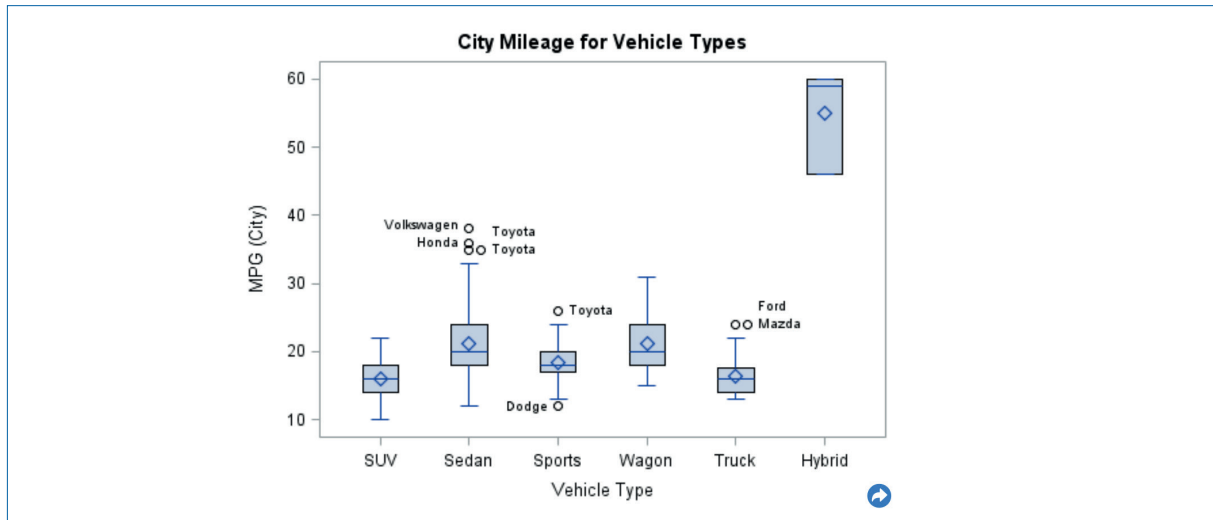


(Own elaboration, simulated data)

In the first case, when the goal is to represent the distribution of a single dataset, the charts that best encode this relationship are simple distribution charts such as the histogram (Figure 22), the frequency polygon (Figure 23), and the box plot (Figure 24).

**Figure 26**

Vehicle distribution by model and miles per gallon ([https://support.sas.com/rnd/datavisualization/gtl/boxplot\\_sect2.htm](https://support.sas.com/rnd/datavisualization/gtl/boxplot_sect2.htm)).



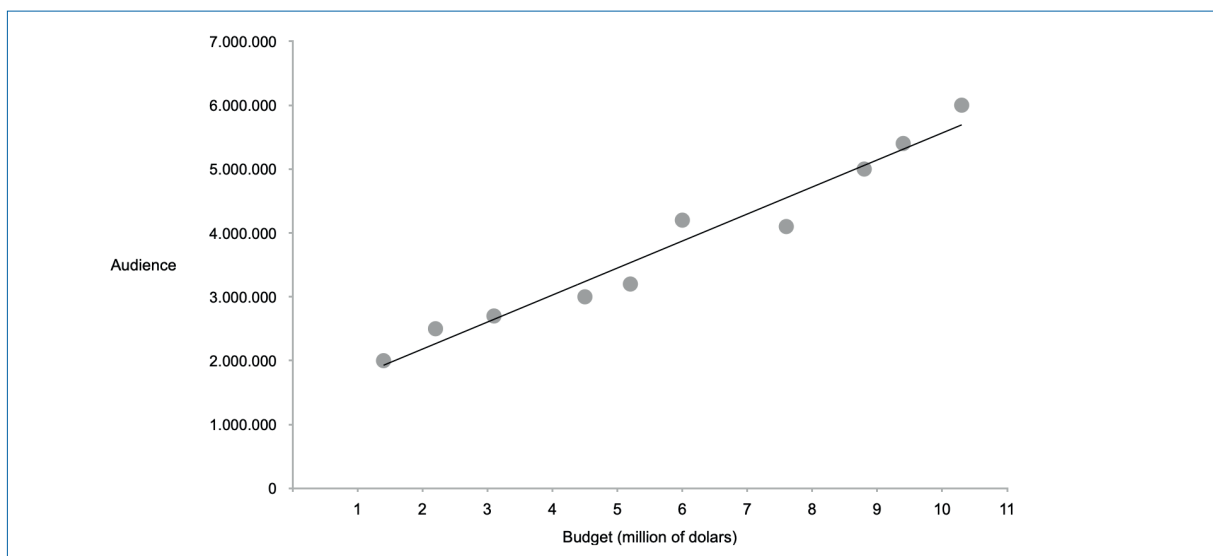
In the second case, when the goal is to simultaneously represent the distributions of two or more datasets, the charts that best encode this relationship are multiple distribution charts, such as combining different frequency polygons in one graph (Figure 25) or overlaying multiple box plots in the same visual layout (Figure 26).

### 3.7. Visualizing correlation

Next, we address the visualization of the final type of semantic content. A correlation relationship, as previously discussed, assigns or connects two quantitative values (each belonging to a distinct quantitative variable) to each item (value or category) of a qualitative variable. This facilitates identifying patterns (strong or weak) in the variation between these two numeric values related to those items.

**Figure 27**

Budget and audience for a film released last week.

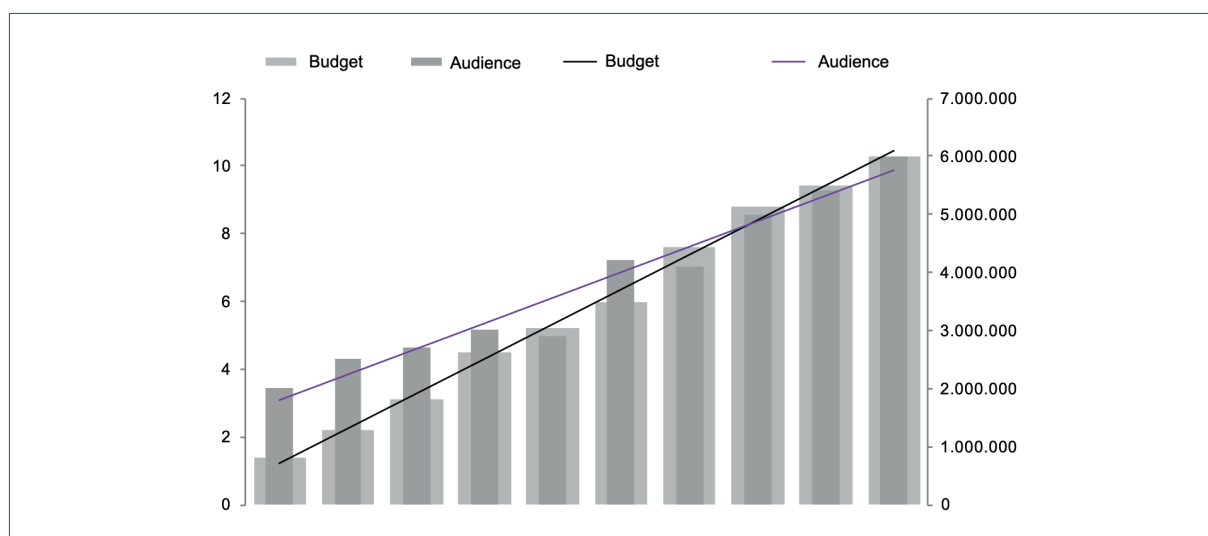


(Own elaboration, simulated data)

A chart that effectively represents this type of semantic content must use visual objects that simultaneously encode the two numeric values and the associated category while expressing whether there is a systematic relationship in which the values of one quantitative variable change relative to the values of the other.

**Figure 28**

*Budget and audience for films released last week.*

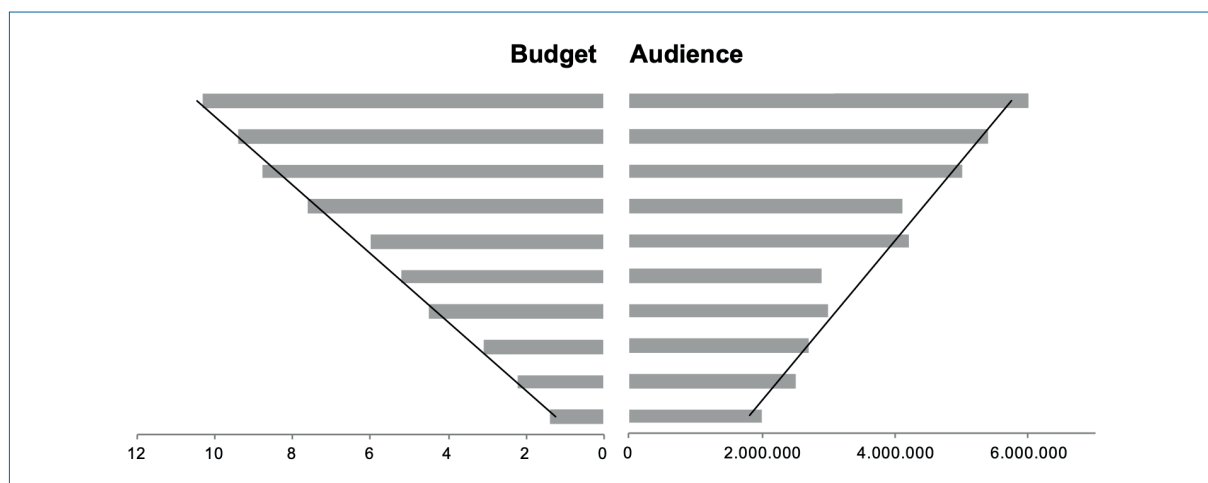


(Own elaboration, simulated data)

The most commonly used visual representation for this type of semantic content is the scatter plot, which can be understood as a specific variant of a dot chart (Figure 27).

**Figure 29**

*Budget and audience for films released last week.*



(Own elaboration, simulated data)

Some authors argue that, particularly in business contexts, scatter plots can be difficult to interpret for users unfamiliar with this type of visualization (Few, 2012). As more intuitive alternatives, when working with small datasets, it is possible to use a correlation column chart (Figure 28) or a paired bar chart (Figure 29).



## 4. Funding

This work is part of the Project “Parameters and strategies to increase the relevance of media and digital communication in society: curation, visualisation and visibility (CUVICOM)”. Grant PID2021-123579OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU.

## 5. References

- Cherven, K. (2015). *Mastering Gephi Network Visualization*. Birmingham: Ed. Packt Publishing Ltd.
- Few, S. (2012). *Show me the numbers*. Oakland: Analytics Press.
- Hearn, D. (2011). *Computer graphics with OpenGL*. Boston: Pearson.
- Hugues, J. F., & Van Dam, A. (2013). *Computer Graphics: Principles and Practice*. Boston: Addison-Wesley.
- Jones, B. (2014). *Communicating data with Tableau: Designing, developing, and delivering data visualizations*. London: Ed. Safari Books Online.
- Khan, A. (2016). *Jumpstart Tableau: A step-by-step guide to better data visualization*. London: Ed. Apress.
- Pérez-Montoro, M. (2022). *Comunicación visual de la información. Qué y cómo podemos narrar con datos*. Río de Janeiro: IBICT-UNESCO.
- Shirley, P. (2009). *Fundamentals of computer graphics*. Natick, MA: AK Peters.
- Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, IEEE Computer Society Press, pp. 336-343.
- Zelazny, G. (2001). *Say it with charts*. New York: MacGraw-Hill.

# Mobilizing news media audiences. Models for the Activation of Relevance and Engagement in Audiences

**Llúcia Castells-Fos**

Universitat Pompeu Fabra, Spain

<https://orcid.org/0000-0003-1124-7941>

**Lluís Codina**

Universitat Pompeu Fabra, Spain

<https://orcid.org/0000-0001-7020-1631>

**Carles Pont-Sorribes**

Universitat Pompeu Fabra, Spain

<https://orcid.org/0000-0003-1135-4245>

---

Castells-Fos, L., Codina, L., & Pont-Sorribes, C. (2025). Mobilizing news media audiences. Models for the Activation of Relevance and Engagement in Audiences. In J. Guallar, M. Vázquez, & A. Ventura-Cisquella (Coords). *Digital communication. Trends and good practices* (pp. 104-118). Ediciones Profesionales de la Información. <https://doi.org/10.3145/cuicom.08.eng>

---

## Abstract

The MAREA (Models for the Activation of Relevance and Engagement in Audiences) framework offers a practical methodology for media outlets to assess and enhance their relevance and engagement capacity, crucial elements for their long-term economic sustainability. This model operationalizes media sustainability through three fundamental dimensions: reputation, visibility, and audience loyalty. MAREA proposes a set of 337 actions and 226 metrics, synthesized into seven main action parameters: scale, user orientation, organizational development, churn prevention, editorial identity, transparency, and institutional recognition. These parameters, alongside clear metrics, enable media organizations, especially smaller ones with limited resources, to identify areas for improvement and develop concrete action plans. The chapter details each parameter, illustrates its application with a hypothetical case, and emphasizes the importance of reinterpreting metrics beyond mere visibility to encompass reputation and audience loyalty, thereby facilitating proactive management and greater resilience in a changing media environment.

## Keywords

News media sustainability; Relevance; Engagement; Audience; MAREA; Framework.

---

## 1. Introduction

The Models for the Activation of Relevance and Engagement in Audiences (MAREA) have been developed to assess media outlets' relevance and engagement capacity. The primary objective is to provide a procedure for evaluating a media outlet's current situation in relation to a set of actions that have proven effective in enhancing relevance and engagement. Secondly, it prioritizes the suggested future improvements should news media managers wish to implement changes.

The MAREA framework is designed to apply to the broadest possible range of media outlets. Its questions can be addressed without financial investment or a major time commitment, relying instead on information that is typically accessible to most media organizations. Larger media outlets usually have analytics teams who fulfill some or all of the goals proposed by this research. Conversely, smaller media outlets with limited resources often lack such capabilities, making MAREA particularly relevant and beneficial for them.

The starting point of this research was the thorough analysis of best practices and expert recommendations from media industry think tanks (Castells-Fos et al., 2025). Throughout these investigations, however, it became clear that both relevance and engagement require more concrete operationalization in order to be activated in the day-to-day work of media professionals (Castells-Fos et al., 2022), which is precisely what MAREA proposes.

To meet this need, we proposed dividing the concept of relevance into two dimensions: (1) reputation and (2) visibility. We also chose to focus on one of the core components of engagement, audience loyalty (dimension 3 for this study), as this is strongly linked to the economic sustainability of media organizations (Castells-Fos et al., 2023b). Collectively, these three dimensions are referred to as the "Considered Sustainability Dimensions."

The three proposed dimensions — reputation, visibility, and audience loyalty — have shown a clear positive impact on revenue generation for media outlets. This suggests that when media organizations create strategies, action plans, and goals specifically focused on enhancing these areas, they can significantly improve long-term sustainability. By prioritizing these dimensions, media outlets can better position themselves to succeed in a competitive environment.

The MAREA recommendations database encompasses actions and metrics — specifically, 337 actions and 226 metrics. Each type of recommendation requires a specific approach and plays a distinct role within the model.

Recommendations in the “action” category allow us to determine the development level of the Considered Sustainability Dimensions — reputation, visibility, and audience loyalty —and establish priorities for designing their improvement plan. Additionally, the metrics will help establish a baseline from which to assess the effectiveness of improvement plans that may be implemented by applying MAREA.

This chapter is a condensed version of the full MAREA report, published independently and in open-access thanks to the Cátedra Ideograma at Pompeu Fabra University (Castells-Fos et al., 2025). The extended version presents the methodology in full detail and offers more comprehensive explanations of each section covered in this chapter.

## 2. The MAREA analytical framework

### 2.1. Actions

The 337 recommendations categorized as actions were synthesized into indicators and then grouped into parameters. Data extraction and subsequent analysis resulted in seven parameters containing 57 indicators. The number of times reflects how often a recommendation appears in the analyzed database, indicating a higher level of expert consensus.

**Table 1**  
 List of Action Parameters.

Action parameters	N. of times mentioned	N. of indicators	Average number of times per indicator
Scale	116	20	5,8
User orientation	64	6	10,7
Organizational development	46	6	7,7
Churn prevention	32	8	4,0
Editorial identity	28	3	9,3
Transparency	26	4	6,5
Institutional recognition	25	10	2,5
Total Mentions	337	57	5,9

Source: Authors’ own elaboration.

The parameters in Table 1 are presented in descending order based on the number of mentions. While the Scale parameter accounts for more than one-third of all recommendations, expert consensus is notably stronger around the user orientation and editorial identity parameters,

with 10.7% and 9.3%, respectively, compared to just 5.8% for scale. This difference may suggest, on the one hand, that the editorial identity parameter contains a more focused and concise set of recommendations. On the other hand, there are parameters — such as scale — where media organizations have a wider range of available options.

## Scale parameter

The scale parameter includes indicators that help enhance a media outlet's visibility in the digital environment. Table 2 presents 20 indicators associated with this parameter, along with brief definitions and the frequency with which each indicator is mentioned in the evidence base. Overall, there are 116 mentions, averaging 5.8 mentions per indicator.

**Table 2**  
*Indicators of the Scale Parameter.*

Scale indicators	N. of times mentioned	Description
Content formats	35	Create a variety of formats to expand the reach of a single piece of journalistic content.
Social media	26	Given that a significant percentage of a news media outlet's audience accesses news through social media rather than its website, it is essential to develop a comprehensive social media strategy to strengthen the outlet's presence across these platforms.
SEO	17	Effective SEO requires a dual approach: Newsroom staff must be trained to implement SEO principles in their writing; additionally, there must be a comprehensive strategy focused on advancing SEO initiatives throughout the entire editorial operation.
Pricing strategy	6	Define a pricing strategy for potential paying users — whether referred to as subscribers, members, or customers — tailored to their position in the conversion funnel, from casual visitors to highly engaged readers ready to support the outlet financially.
Experimentation	5	Experimentation does not require a media lab. Any outlet can test new services, content, page designs, multimedia formats, or even subscription offers. The goal is to foster continuous improvement and better audience understanding.
Syndication	3	Entering into agreements with other outlets to republish content can expand reach and help grow audience size.
Acquisition	3	To enhance traffic, outlets should implement targeted campaigns in conjunction with their ongoing social media activity. This involves designing promotional initiatives that align with the outlet's brand, values, and specific objectives.
Calls to action	3	Optimal placement of calls to action on the website — ensuring good visibility without relying on intrusive banners. Doing it regularly helps to communicate the outlet's need for financial support from its audience, fostering a sense of community involvement and shared responsibility toward the outlet's sustainability.
Branded content	2	A branded content strategy focuses on creating advertiser-funded content specifically designed to promote a brand effectively. This approach diversifies the advertising portfolio, thereby increasing advertising revenue streams.
Referral program	2	These programs can generate snowball effects, increasing subscribers and enhancing content visibility. For example, the program might allow subscribers to share a limited number of articles with friends and family.
Agrupación de audiencias	2	Partner with other outlets to create more attractive audiences for advertisers. This strategy can significantly boost advertising revenue by making the outlet's platform more appealing to a wider range of advertisers.
Influencers	2	Partner with influencers and well-known public figures to harness the power of the influencer phenomenon. This presents a valuable opportunity for media outlets to expand their reach and engage with new audiences effectively.
First-party data	2	Collect information directly from your audience — such as reading habits, preferences, or subscription history — rather than relying on third parties. This data not only helps outlets better understand their readers and improve content or services, but also allows them to offer advertisers more targeted and valuable opportunities for direct sales.

Scale indicators	N. of times mentioned	Description
Author's profiles	2	Encourage the newsroom team to maintain an active presence on social media. Outlets should define clear policies for journalist activity on social platforms, potentially recommending active engagement while agreeing on appropriate forms and procedures.
Multilingual publishing	1	Publish content in more than one language to reach broader or more diverse audiences. Adapting to the linguistic realities of different communities — whether local, national, or international — can significantly expand an outlet's visibility and impact.
Accessibility	1	Accessibility refers to designing a website that is accessible to people with disabilities. Subtitling audiovisual materials is especially important. It also increases visibility by making such content searchable by web crawlers.
Automation	1	Use automated tools to produce certain types of news —especially those based on data, like weather, sports results, or financial updates. Automation can help increase content output and free up time for journalists to focus on more complex stories. When used, it should be transparent so that audiences understand how the content was generated.
Content volume	1	Assess and optimize the volume of content produced to align with the outlet's visibility and audience growth goals. While more content isn't always better, increasing output — when starting from a low baseline — can help improve reach and relevance.
Free trial	1	In the case of paywalls, news media outlets should consider free trial periods to attract potential subscribers.
Content distribution	1	Establish partnerships with companies or organizations (not necessarily media outlets) to distribute content through additional channels. These agreements can help expand the outlet's reach and attract new audiences beyond its core platform.
Total mentions	116	

Source: Authors' own elaboration.

### User orientation parameter

The user orientation parameter brings together indicators aimed at focusing the media outlet's activities on the needs and behaviours of its users. Table 4 displays the six indicators related to this parameter, along with the frequency of each indicator's occurrence in the evidence base. In total, there are 64 mentions averaging 10.7 mentions per indicator.

**Table 3**  
*Indicators of the user orientation parameter.*

User orientation indicators	N. of times mentioned	Description
Audience participation	24	Encourage audience participation through various channels, whether digital or traditional.
Audience understanding	17	Gather audience feedback through multiple channels to monitor user behaviour and analyze their needs and interests.
Conversion funnel optimization	9	Establish a conversion funnel (from free users to paying customers) as the foundation for an optimization plan focused on enhancing the user's web experience. The funnel primarily involves analyzing browsing behaviour.
Personalization	8	Personalize communication with users based on their monitored behaviour.
Tone of voice	3	Use clear, relatable, and appropriate language when communicating with the target audience. For instance, some outlets practice explanatory journalism, which requires a greater depth of context and a thoughtful approach to language.
Comment policy	3	A clear policy for responding to comments is recommended to those who maintain their comment sections and need to establish clear guidelines to protect the outlet's brand.
Total mentions	64	

Source: Authors' own elaboration.

## Organizational development parameter

When confronted with financial sustainability issues, experts frequently emphasize the critical role played by organizational factors. The organizational development parameter brings together indicators related to the human and technological resources required within a media organization. As detailed in Table 6, there are six key indicators linked to this parameter, with a total of 46 instances, averaging 7.7 times per indicator.

**Table 4**  
*Indicators of the organizational development parameter*

Organizational development indicators	N. of times mentioned	Description
Newsroom technology	24	Equip the organization with the right digital tools to support editorial work across teams, with a clear focus on serving the audience efficiently and collaboratively.
Support roles	12	Build a multidisciplinary team beyond the newsroom, including roles such as web design, tech support, audience insights, membership management, ad sales, community engagement, customer service, and corporate communications — key functions that complement and enhance the editorial work.
Internal training	7	Provide ongoing training and structured meetings to enhance day-to-day operations, facilitate the effective use of available tools by teams, share best practices, and align everyone around shared objectives.
Advertiser reporting	1	Develop standardized reporting formats tailored to the needs of advertisers.
Talent retention	1	Create strategies to retain top editorial talent. Media organizations invest heavily in their teams and must ensure that key contributors remain and grow within the company.
System integration	1	Connect internal management systems — especially those handling user or subscriber data — to streamline workflows, reduce duplication, and strengthen collaboration across departments.
Total mentions	46	

Source: Authors' own elaboration.

## Churn prevention parameter

There is a broad consensus among experts that acquiring new subscribers — primarily through special campaigns or heavily discounted offers, sometimes at less than 50% of the regular price — is relatively easy. In contrast, subscriber churn poses a much greater challenge. It is no surprise, then, that reports from think tanks and expert recommendations emphasize the need to minimize subscriber loss, which can stem from a wide range of causes, as illustrated by the indicators discussed below.

The churn prevention parameter combines indicators aimed at increasing the retention rate of paying users or, conversely, reducing their likelihood of cancellation. Table 8 lists the eight indicators included under this parameter, which appear a total of 32 times in the evidence base, averaging 4.0 mentions per indicator.



**Table 5**  
*Indicators of the churn prevention parameter.*

Churn prevention indicators	N. of times mentioned	Description
Subscriber incentives	10	It is essential to highlight the advantages of becoming a subscriber. Media outlets should provide exclusive content to subscribers, making the benefits of subscribing clear. Some outlets offer non-subscribers partial access, while others provide additional features exclusively for paying subscribers.
Management of expired payment methods	6	Establish a system to manage expired payment methods, such as outdated credit cards used for subscription payments. Experts insist that a significant percentage of cancellations can be attributed to issues related to expired payment information.
Onboarding process	6	Establish a structured onboarding process for new subscribers. In this context, “onboarding” refers to the process by which a subscriber becomes familiar with the outlet’s services.
Cancellation risk detection	4	Define a usage threshold below which there is a risk of cancellation. Data shows that subscribers who are inactive for several days or exhibit minimal engagement, such as not reading articles or having low newsletter open rates, are at a higher risk of cancellation. This information can help trigger targeted re-engagement efforts.
Long-term subscription incentives	2	Offer discounts for long-term subscriptions. Securing subscribers for a year or more is especially important, as it reflects a higher level of commitment and dedication. Discount strategies can help convert monthly subscribers into annual subscribers.
Cancellation feedback analysis	2	Analyze the reasons behind cancellations. After a subscriber cancels, collecting their feedback —typically through a brief survey— can provide valuable insights into the causes of churn.
Reactivation campaigns	1	Implement re-engagement campaigns targeting former subscribers. Non-intrusive campaigns can encourage them to resume their subscription.
Subscriber fatigue mitigation	1	Develop curated content bundles to prevent subscriber fatigue —a phenomenon identified by experts as a key driver of cancellations.
Total mentions	32	

Source: Authors’ own elaboration.

### *Editorial identity parameter*

The parameter referred to as Editorial Identity encompasses a set of indicators related to how a media outlet defines and communicates its brand positioning in comparison to its competitors. It includes three indicators, which appear a total of 28 times in the evidence base, averaging 9.3 mentions per indicator.

**Table 6**  
*Indicators of the editorial identity parameter*

Editorial identity indicators	No. of times mentioned	Description
Positioning	15	Define a precise editorial positioning that distinguishes the outlet from competitors and establishes a unique relationship with its audience.
Editorial focus	8	Choose a distinctive mix of journalistic formats — such as news reporting, opinion pieces, or in-depth features — as part of what defines the outlet’s editorial personality.
Topic specialization	5	Build identity around specific themes or subject areas. While large media organizations may cover a wide range of topics, smaller or mission-driven outlets often stand out by focusing on niche or underrepresented content.
Total mentions	28	

Source: Authors’ own elaboration.

## Transparency parameter

Growing phenomena such as declining trust in the media — and related trends like news avoidance — have led experts to place strong emphasis on transparency. Ethical principles also play a significant role in this demand. This parameter combines indicators that promote transparency, thereby building trust with the audience. As shown in the table below, it includes four indicators with a total of 26 mentions, averaging 6.5 per indicator.

**Table 7**  
*Indicators of the transparency parameter.*

Transparency Indicators	N. of times mentioned	Description
Editorial standards	15	Establish clear editorial standards or an ethical code to guide professional practices and journalistic quality.
Accountability	5	Be transparent about how the organization is managed and its funding sources. Transparency is crucial for building trust, particularly in loyalty or membership-based models. It could involve publishing corporate or financial information.
Fact-checking	4	Implement consistent, well-defined, and rigorous fact-checking processes within the newsroom.
Information vs. opinion and advertising	2	Clearly distinguish between sponsored content (such as branded content or ads), opinion pieces, and factual news reporting.
Total mentions	26	

Source: Authors' own elaboration.

## Institutional recognition parameter

This parameter also brings together a set of indicators that, according to the experts consulted, can contribute to strengthening public recognition when implemented. As reflected in the following table, the 10 indicators that make up this parameter appear a total of 25 times in the evidence base, for an average of 2.5 mentions per indicator.

**Table 8**  
*Indicators of the institutional recognition parameter.*

Institutional recognition indicators	N. of times mentioned	Description
Public Presence	6	Develop a media strategy for editors and newsroom leadership for their public appearances.
Contributors	4	Invite respected contributors to collaborate with the outlet, enhancing its editorial identity.
Explanatory Journalism	3	Publish explanatory content that helps audiences understand key issues. This type of journalism focuses on breaking down complex topics and presenting them in a way that is accessible and engaging for the audience.
Collaborative Projects	2	Collaborate on joint reporting projects with other media outlets. By pooling resources and expertise, media organizations can tackle larger stories and share diverse perspectives, resulting in richer narratives that benefit the public and elevate the quality of journalism.
Culture of Innovation	2	Foster a culture of innovation within the newsroom to develop better products and demonstrate forward-thinking values to audiences and stakeholders.

Institutional recognition indicators	N. of times mentioned	Description
Quality vs. Quantity Balance	2	Define a strategy that balances content volume with high editorial standards. Prioritizing quality over sheer quantity ensures that the journalism remains credible and trustworthy, fostering audience loyalty and enhancing the overall reputation of the outlet.
Exclusives	2	Establish a straightforward approach to exclusives in areas where the outlet has a strong voice or expertise, reinforcing its editorial identity.
Impartiality and Balance	2	Avoid overly partisan coverage where possible. Fairness and neutrality should be prioritized.
Sponsor Alignment	1	Select sponsors whose values align with the outlet's identity and mission. Clear sponsorship policies help preserve editorial independence and audience trust.
Journalism Awards	1	Submit work for journalism awards to increase visibility and strengthen the outlet's reputation.
Total Mentions	25	

Source: Authors' own elaboration.

### Example of application

The following example illustrates how the MAREA matrix can be applied to a hypothetical case involving a simulated media outlet. This comparison highlights the outcomes achieved by the outlet against an ideal scenario that assumes the full implementation of all MAREA recommendations. A media outlet may opt to evaluate all indicators or select specific ones based on its unique context. For the sake of simplicity, this simulation presumes that the outlet has addressed every indicator. The focal parameter for this example is organizational development.

The table below presents the results. An exemplary media outlet that fulfils all requirements would obtain a maximum score of 39.5, while the simulated case achieves a score of 29.9. Figure 1 visually represents this performance, indicating that the simulated outlet has reached 76% of the ideal benchmark.

**Table 9**

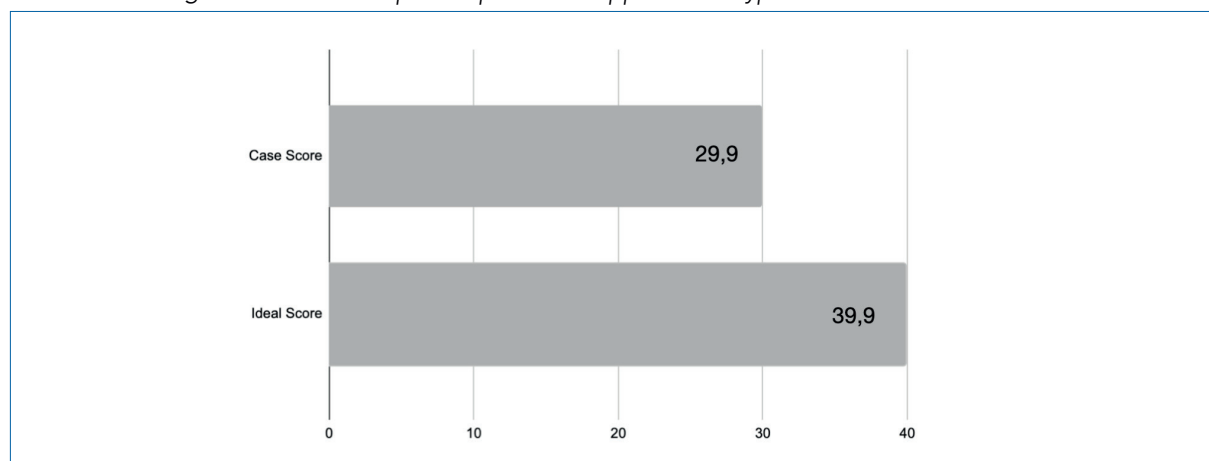
*Application of the 'Organizational Development' parameter to a hypothetical case*

Organizational development indicators	Score		FdF			Results	
	Ideal	Case	FdF1	FdF2	FdF3	Ideal	Case
Newsroom Technology	1	1	4,0	3	2	24,0	24,0
Support Roles	1	0	1,6	3	2	9,4	0,0
Internal Training	1	1	0,9	3	2	5,5	5,5
Advertiser Reporting	1	1	0,1	1	1	0,1	0,1
Talent Retention	1	0	0,1	1	2	0,3	0,0
System Integration	1	1	0,1	1	2	0,3	0,3
Total						39,5	29,9

Source: Authors' own elaboration.

**Figure 1**

Result of the 'Organizational Development' parameter applied to a hypothetical case



Source: Authors' own elaboration.

## 2.1. Metrics

MAREA's core contribution lies in its ability to translate sustainability into actionable steps within newsrooms. Its goal is to activate, in a structured and prioritized way, a set of concrete measures that help build more resilient media. Metrics play a supporting role in this process.

As explained in Chapter 2 of the MAREA Guide – General Objectives and Procedures, the approach draws on simple, concrete indicators that reflect a baseline consensus among the experts consulted. These indicators account for 53% of the total metric-based recommendations in the database.

**Table 10**

List of Simple, Concrete Metric Indicators

Metric Indicator	N. of times mentioned	Description
Reach (unique visitors)	15	Total number of individual users who access your site over a defined period — your core visibility metric.
Active supporters (paying users)	13	People or organizations who financially contribute to your outlet via subscriptions, memberships, or donations.
Audience retention (returning users)	13	Users who return to your site after an initial visit — critical for loyalty and long-term growth.
Revenue generation	10	Total income from all sources, including ads, subscriptions, events, or grants.
Social audience (followers)	8	Total number of followers across major social media platforms — key for brand visibility.
Social engagement	8	Total interactions (likes, shares, comments) across social channels, indicating content resonance.
Email list growth (newsletter subscribers)	7	Total newsletter opt-ins — a powerful channel for direct engagement and conversion.
Content consumption (page views)	6	The volume of pages viewed on your site — a sign of content traction.
Engagement depth (session duration)	5	Average time spent per visit, revealing how compelling your content experience is.
Brand longevity (years active)	4	The number of years the outlet has operated — can build credibility and trust.

Metric Indicator	N. of times mentioned	Description
Email engagement (open rate)	4	Share of emails opened out of total sent — an indicator of content relevance and list health.
Bounce rate	3	The percentage of visits where users didn't engage — useful for diagnosing weak points in UX or targeting.
Referral traffic	3	Visits coming from other websites —shows how well your content is shared or syndicated.
Email click-throughs	3	Number of clicks within newsletters —vital for funnel movement and conversion.
Brand credibility (awards received)	2	Recognitions and awards that enhance your outlet's authority and positioning.
Supporter conversion rate	2	Percentage of unique users who become paying supporters — your monetization efficiency metric.
Subscriber churn rate	2	The rate at which subscribers cancel —essential for assessing product satisfaction and retention efforts.
Total Mentions	121	

Source: Authors' own elaboration.

### 3. Conclusions

MAREA provides a framework to activate the core dimensions of media sustainability: reputation, visibility, and audience loyalty. Easy to implement and straightforward to interpret, it provides practical guidance for media outlets seeking to develop a roadmap for improvement. The seven action parameters in MAREA range from strategic (such as editorial identity) to tactical (like churn prevention).

Scale is the parameter that receives the most mentions and is characterized by a strong technical focus. Closely tied to visibility, its effectiveness, however, is amplified when anchored in the reputation of the media outlet as indicators such as SEO, syndication, and acquisition suggested. This category encompasses the highest number of indicators within the MAREA framework, mainly due to extensive expert input; however, not all indicators are consistently aligned. Notably, six out of the 20 indicators in this category are referenced only once.

User orientation shows the highest level of consensus, with an average of 10.7 mentions per indicator. It requires a deep understanding of the audience, including their expectations, needs, and behaviours, to capture their interest, ensure recurrence, encourage participation, and ultimately convert them into paying users. These practices generate a wealth of data that provides feedback to inform future decisions.

Organizational development is a strategic parameter that focuses on building the structural foundation needed for long-term sustainability. It aims to ensure that media outlets are equipped with the right personnel and tools to enhance their capacity and streamline operations.

Churn prevention is tactical but essential. In competitive markets saturated with alternatives — and with subscription fatigue on the rise — this parameter is crucial to maintaining steady revenue from readers.

Editorial identity and transparency are strategic and essential for news organizations. The former influences content direction, while the latter builds trust through responsible content creation processes.

Institutional recognition ranks second in the number of indicators (10), though the average consensus is relatively low (2.5 mentions per indicator, compared to a database average of 5.9). This low ratio reflects the wide range of ideas experts have proposed for strengthening reputation — a goal often framed in terms such as *impact*, *influence*, or *trust*.

As for metrics, MAREA offers a streamlined approach to help outlets establish a baseline set of indicators with minimal effort. These metrics are most valuable when used as a diagnostic tool to inform the development of an improvement plan. Experts agree most on size-related metrics, as both academia and industry have finally accepted the need to work with the imperfect but available information these metrics provide. The key is to reinterpret the metrics not solely in terms of visibility (their most apparent application), but also in terms of reputation and audience loyalty.

This reinterpretation has created a cross-metric utility: among the 24 indicators listed, most evaluate two or even all three dimensions. Specifically, 16 indicators relate to reputation, 12 to visibility, and 17 to loyalty.

Although MAREA is based on studies conducted between 2022 and 2024, most of the indicators are expected to remain relevant in the medium term. Besides, its most significant contribution is meant to be both methodological and conceptual, providing a flexible framework that professionals and researchers alike can adapt and expand. The goal is to develop essential managerial skills for media practitioners, enabling them to succeed in a fast-changing environment, shaping the future of journalism and ensuring its sustainability.

## 4. Funding

This work is part of the Project “Parameters and strategies to increase the relevance of media and digital communication in society: curation, visualisation and visibility (CUVICOM)”. Grant PID2021-123579OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU.

## 5. References

This section includes the references, both cited and consulted, that have contributed to the development of MAREA.

ADEPA. (2022). *Industria periodística 2021. Anuario 2021*. <https://adepa.org.ar/wp-content/uploads/2021/12/anuario-adepa-2021.pdf>

Amoedo-Casais, A., Vara-Miguel, A., Negrodo-Bruna, S., Moreno-Moreno, E., & Kaufmann-Argueta, J. (2021). *Digital News Report España*. <https://dadun.unav.edu/handle/10171/62862>

Carlson, M. (2018). Confronting measurable journalism. *Digital Journalism*, 6(4), 406-417. <https://doi.org/10.1080/21670811.2018.1445003>

Castells-Fos, L., Codina, L., & Pont-Sorribes, C. (2025). Modelos de activación de la relevancia y el engagement de audiencias en medios de comunicación. UPF. Cátedra Ideograma. <https://doi.org/10.31009/com.2025.01>

Castells-Fos, L., Lopezosa, C., Pont-Sorribes, C., & Codina, L. (2023). Reputación, visibilidad y fidelización de audiencias de medios: una propuesta operativa desde la academia. *Estudios Sobre el Mensaje Periodístico*, 29(2), 419–430. <https://doi.org/10.5209/esmp.85655>

- Castells-Fos, L., Pont-Sorribes, C., & Codina, L. (2023). Decoding news media relevance and engagement through reputation, visibility and audience loyalty: A scoping review. *Journalism Practice*, 19(5), 1154-1173. <https://doi.org/10.1080/17512786.2023.2239201>
- Castells-Fos, L., Pont-Sorribes, C., & Codina, L. (2022). La sostenibilidad de los medios a través de los conceptos de *engagement* y relevancia: scoping review. *Doxa Comunicación*, 35, 19-38. <https://doi.org/10.31921/doxacom.n35a1627>
- Clear, J. (s.f.). *All models are wrong, some are useful*. James Clear. <https://jamesclear.com/all-models-are-wrong>
- Day, B., & Wills, F. (2020). 2020 Advertiser assessment of local media digital advertising.
- Del Pozo, M. (2022). Netflix se lanza a la publicidad. *Expansión*. <https://www.expansion.com/blogs/peon-de-dama/2022/10/23/netflix-se-lanza-a-la-publicidad.html>
- Escolar, I. (2021, July 3). Las cuentas de 2020: los socios y socias salvan elDiario.es. *ElDiario.es*. [https://www.eldiario.es/escolar/cuentas-socios-socias-salvan-eldiario\\_132\\_8102443.html](https://www.eldiario.es/escolar/cuentas-socios-socias-salvan-eldiario_132_8102443.html)
- Escolar, I. (2020). Un periódico que se debe a su comunidad: las cuentas de eldiario.es en 2019. *ElDiario.es*. [https://www.eldiario.es/escolar/cuentas\\_132\\_6071575.html](https://www.eldiario.es/escolar/cuentas_132_6071575.html)
- Espada, A. (2021). Impact of the COVID-19 Pandemic on Media sustainability in Latin America. [https://unesdoc.unesco.org/ark:/48223/pf0000377631\\_eng](https://unesdoc.unesco.org/ark:/48223/pf0000377631_eng)
- Evolok. (2021). The paywall revolution goes global. <https://pressgazette.co.uk/subscription-news-media-must-re-think-model-for-creating-new-revenues/>
- Ferrer-Conill, R., & Tandoc, E. C. (2018). The audience-oriented editor: Making sense of the audience in the newsroom. *Digital Journalism*, 6(4), 436-453. <https://doi.org/10.1080/21670811.2018.1440972>
- Forman, C. I. (2021, October 12). Solutions to America's Local Journalism Crisis: Consolidated Literature Review. <https://shorensteincenter.org/solutions-americas-local-journalism-crisis-consolidated-literature-review>
- Graves, A. L., & Kelly, J. (2010). *Confusion online: Faulty metrics and the future of digital journalism*. Tow Center for Digital Journalism. <https://search.issuelab.org/resources/15315/15315.pdf>
- Gundlach, H., & Hofmann, U. (2021). Information search, behavioural economics, and relevance decisions in the online media industry: how strongly do the algorithms of intermediaries influence the relevance evaluation of information? *Journal of Media Business Studies*, 18(3), 179-198. <https://doi.org/10.1080/16522354.2020.1854602>
- Guzik, S. (2022). *The global survey on journalism's futures*. Future Today Institute. [https://www.linkedin.com/posts/samguzik\\_the-global-survey-on-journalisms-futures-activity-6912124871074299904-lmke](https://www.linkedin.com/posts/samguzik_the-global-survey-on-journalisms-futures-activity-6912124871074299904-lmke)
- Iastrebnier, M., Warner, J., & Best, J. (2021). Punto de Inflexión.
- Kovach, B., & Rosenstiel, T. (2001). *The elements of journalism* (3rd ed.). Three Rivers Press.
- Krebs, I., Bachmann, P., Siegert, G., Schwab, R., & Willi, R. (2021). Non-journalistic competitors of news media brands on Google and YouTube: From solid competition to a liquid media market. *Journal of Media Business Studies*, 18(1), 27-44. <https://doi.org/10.1080/16522354.2020.1832746>



- Li, A. (2021). How customer service can build trust and engagement with audiences. <https://www.americanpressinstitute.org/publications/reports/strategy-studies/how-customer-service-can-build-trust-and-engagement-with-audiences/>
- López-García, G. (2015). *Periodismo digital: redes audiencias y modelos de negocio*. Comunicación Social.
- López-Linares, C. E. (2021). *Innovación y periodismo local en América Latina*. Fundación Gabo. <https://fundaciongabo.org/es/recursos/publicaciones/fundacion-gabo-y-google-news-initiative-lanzan-libro-digital-innovacion-y>
- Lopezosa, C., Codina, L., & Freixa, P. (2022). *ATLAS.ti para entrevistas semiestructuradas : guía de uso para un análisis cualitativo eficaz*. DigiDoc Research Group | Pompeu Fabra University, 30. <http://repositori.upf.edu/handle/10230/52848>
- Matsa, K. E., & Naseer, S. (n.d.). News platform fact sheet. <https://www.pewresearch.org/journalism/fact-sheet/news-platform-fact-sheet/>
- McKisson, I. & Pallack, B. (2021). What types of messages, pricing, and perks motivate #ThisIsTucson super users to become paying members.
- Nelson, J. L., & Tandoc, E. C. (2019). Doing «well» or doing «good»: What audience analytics reveal about journalism's competing goals. *Journalism Studies*, 20(13), 1960-1976. <https://doi.org/10.1080/1461670X.2018.1547122>
- Newman N. (2019). *Journalism, media, and technology trends and predictions 2019*. Reuters Institute. [https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2019-01/Newman\\_Predictions\\_2019\\_FINAL\\_2.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2019-01/Newman_Predictions_2019_FINAL_2.pdf)
- Newman, N. (2021). Digital news report: Journalism, media, and technology trends and predictions 2021. Reuters Institute. [https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2021-01/Newman\\_Predictions\\_2021\\_FINAL.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2021-01/Newman_Predictions_2021_FINAL.pdf)
- Newman, N., Fletcher, R., Robertson, C. T., Eddy, K., & Nielsen, R. K. (2022). *Digital news report 2022*. Reuters Institute. [https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2022-06/Digital\\_News-Report\\_2022.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2022-06/Digital_News-Report_2022.pdf)
- Newman, N., Fletcher, R., Eddy, K., Robertson, C. T., & Nielsen, R. K. (2023). *Digital news report 2023*. Reuters Institute. [https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2023-06/Digital\\_News\\_Report\\_2023.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2023-06/Digital_News_Report_2023.pdf)
- Oliver, L. (n.d.). *Private gatekeepers: Encrypted messaging Apps and news audiences*. [https://www.cima.ned.org/wp-content/uploads/2022/02/CIMA\\_EMA-Report\\_web\\_150ppi.pdf](https://www.cima.ned.org/wp-content/uploads/2022/02/CIMA_EMA-Report_web_150ppi.pdf)
- Piechota, G. (2021). The benefits and risks of media data democratisation.
- Pope, K. (2017). Is there a business model for real journalism? *Nation*, 304(9).
- Powers, E. (2018). Selecting metrics, reflecting norms: How journalists in local newsrooms define, measure, and discuss impact. *Digital Journalism*, 6(4), 454-471. <https://doi.org/10.1080/21670811.2018.1445002>
- Puente-Rivera, I., Campos-Freire, F & López-García, X (Coords.). (2018). *Periodismo con futuro*. Media XXI.
- PWC. (2021). *Entertainment and media outlook 2021-2025*. PricewaterhouseCoopers. <https://www.pwc.es/es/entretenimiento-medios/entertainment-media-outlook-espana-2021-2025.html>

- Radcliffe, D., & Wallace, R. (2021, October 7). Life at local newspapers in a turbulent era. *Columbia Journalism Review*. [https://www.cjr.org/tow\\_center\\_reports/life-at-local-newspapers-in-a-turbulent-era-findings-from-a-survey-of-more-than-300-newsroom-employees-in-the-united-states.php](https://www.cjr.org/tow_center_reports/life-at-local-newspapers-in-a-turbulent-era-findings-from-a-survey-of-more-than-300-newsroom-employees-in-the-united-states.php)
- Roseman, E., McLellan, M., & Holcomb, J. (2021). The state of nonprofit news.
- Schulte, A. (2021). *A quemar las naves. Guía para activar el cambio cultural en medios de comunicación*. WAN-IFRA. <https://wan-ifra.org/wp-content/uploads/2021/04/WAN-IFRA-Cultural-Change-Espanol-New.pdf>
- Sonderman, J., & Vargo, G. (2021). What news publishers do to retain subscribers. <https://www.americanpressinstitute.org/publications/reports/survey-research/what-news-publishers-do-to-retain-subscribers/single-page/#:~:text=Almost all publishers said they,welcome emails to new subscribers>
- Stocking, G., & Khuzam, M. (2021). *Digital news fact sheet*. Pew Research Center. <https://www.pewresearch.org/journalism/fact-sheet/digital-news/>
- Stroud, N. J. (2017). Helping newsrooms work toward their democratic and business objectives. In P. J. Boczkowski, & C. W. Anderson (Eds.), *Remaking the news: Essays on the future of journalism scholarship in the digital age*. MIT Press.
- The Economist. (2023, December 14). Can you have a healthy democracy without a common set of facts? *The Economist*. <https://www.economist.com/leaders/2023/12/14/can-you-have-a-healthy-democracy-without-a-common-set-of-facts>
- The Economist. (2023, December 14). How American journalism lets down readers and voters. *The Economist*. <https://www.economist.com/united-states/2023/12/14/how-american-journalism-lets-down-readers-and-voters>
- Vallés, M. S. (2003). *Entrevistas cualitativas*. Centro de Investigaciones Sociológicas. <https://doi.org/978-84-7476-342-3>
- Villafañe, J., Ortiz-de-Guinea-Ayala, Y., & Martín-Sáez, J.-L. (2020). Reputación de los medios de comunicación españoles. *Profesional de la Información*, 29(4). <https://doi.org/10.3145/epi.2020.jul.07>

# Factors for enhancing visibility in digital repositories: Metadata quality, interoperability standards, persistent identifiers, and SEO-GEO optimization

**Danilo Reyes-Lillo**

Universitat Pompeu Fabra, Spain

<https://orcid.org/0000-0002-0141-8324>

**Cristòfol Rovira**

Universitat Pompeu Fabra, Spain

<https://orcid.org/0000-0002-6463-3216>

**Alejandro Morales-Vargas**

Universidad de Chile, Chile

<https://orcid.org/0000-0002-5681-8683>

---

Reyes-Lillo, D., Rovira, C., & Morales-Vargas, A. (2025). Factors for enhancing visibility in digital repositories: Metadata quality, interoperability standards, persistent identifiers, and SEO-GEO optimization. In J. Guallar, M. Vázquez, & A. Ventura-Cisquella (Coords). *Digital communication. Trends and good practices* (pp. 119-133). Ediciones Profesionales de la Información. <https://doi.org/10.3145/cuvicom.09.eng>

---

## Abstract

In the context of open science and scholarly communication, enhancing the visibility of digital repositories is essential for maximizing the reach, impact, and discoverability of the content they host. This chapter explores five key strategies to improve repository visibility: optimizing metadata quality, enabling interoperability protocols, adopting persistent identifiers (PIDs), implementing search engine optimization (SEO) strategies, and embracing generative engine optimization (GEO). Through detailed analysis and practical recommendations, this work highlights how standardized metadata, controlled vocabularies, and persistent identifiers, such as DOI, Handle, and ARK, contribute to enhancing visibility. This chapter also emphasizes the importance of aligning repositories with evolving web technologies and AI-driven engines to ensure content remains accessible, traceable, and integrated into users' search experience.

## Keywords

Digital repositories; Visibility; Metadata quality; Interoperability; Persistent identifiers; Search Engine Optimization; Generative Engine Optimization.

---

## 1. Introduction

In the era of open access and open science, digital repositories play a crucial role in disseminating knowledge. Enhancing their visibility not only broadens the reach of deposited content but also strengthens institutional impact and contributes to the democratization of information access. Various studies have shown that open-access publications, particularly those hosted in institutional repositories, tend to receive more citations and are more accessible than those restricted by paywalls (Piwowar et al., 2018; Swan, 2010).

Moreover, increased visibility in academic search engines, such as Google Scholar, and databases like OpenAIRE and CORE, enables repositories to comply with open-access mandates set by national and international funders (UNESCO, 2021). Strategies such as the proper implementation of standardized metadata, the use of persistent identifiers (such as DOIs and ORCID) and interoperability with other systems through protocols like OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) are key to achieving greater exposure (OpenAIRE, 2020).

Therefore, actively working to optimize repository visibility is not merely a technical concern (Reyes-Lillo et al., 2025) but an institutional strategy to ensure that intellectual output fulfills its ultimate purpose: to be discovered, used, and cited by academic communities and society at large.

The following section analyzes five optimization techniques that can be employed to improve the visibility of digital repository content:

1. Metadata Quality Optimization.
2. Enabling Interoperability Protocols.
3. Adoption of Persistent Identifiers.
4. SEO Optimization of the Repository.
5. Generative Engine Optimization in Repositories.

## 2. Metadata quality optimization

Metadata standardization is crucial for ensuring the interoperability, visibility, access, and reuse of deposited content. A metadata strategy must consider not only technical aspects but also organizational and policy-related dimensions of the repository.

While the metadata model is often closely tied to the platform on which the repository is built, the schema must be adapted to local needs without compromising compatibility with international standards.

Table 1 presents a comparison of several digital repository software platforms and their underlying metadata schemas:

**Table 1**

*Digital repository software and their base metadata schemas.*

Software	Base Metadata Schema	Additional information
DSpace	Qualified Dublin Core <a href="https://github.com/DSpace/DSpace/blob/main/dspace/config/registries/dublin-core-types.xml">https://github.com/DSpace/DSpace/blob/main/dspace/config/registries/dublin-core-types.xml</a>	Uses a custom profile of Qualified Dublin Core; supports extensions like METS/MODS. From DSpace 7 onward, multiple schemas are supported.
Fedora	No predefined schema <a href="https://wiki.lyrasis.org/display/FEDORA6x/Data+Modeling">https://wiki.lyrasis.org/display/FEDORA6x/Data+Modeling</a>	Employs RDF/Linked Data models; schema depends on implementation (MODS, DC, PREMIS, etc.).
EPrints	Configurable and extensible metadata schema <a href="https://wiki.eprints.org/w/Metadata#Metadata_Field_Types">https://wiki.eprints.org/w/Metadata#Metadata_Field_Types</a>	Schema based on fields defined by the repository administrator. Can interoperate with other schemas via import/export.
InvenioRDM	JSON structure aligned with DataCite <a href="https://inveniordm.docs.cern.ch/reference/metadata">https://inveniordm.docs.cern.ch/reference/metadata</a>	Schema conforms to DataCite's Metadata Schema v4.x with minor additions and modifications.
TIND	MARC21 <a href="https://www.tind.io">https://www.tind.io</a>	Based on MARC21 due to its foundation in INVENIO (developed by CERN). Offers various modules and can adapt to other formats.
Digital Commons	No specified schema; mappable to Dublin Core <a href="https://digitalcommons.elsevier.com/en_US/organization-content-planning/metadata-options-in-digital-commons">https://digitalcommons.elsevier.com/en_US/organization-content-planning/metadata-options-in-digital-commons</a>	Declares a flexible metadata schema.
DataVerse	Uses various standard-compliant metadata schemas <a href="https://guides.dataverse.org/en/latest/user/appendix.html">https://guides.dataverse.org/en/latest/user/appendix.html</a>	Ensures interoperability and preservation through schemas like DDI, DataCite, Dublin Core, ISA-Tab, and VOResource, enabling structured export.

To initiate a metadata optimization strategy, it is recommended to begin with an initial assessment that includes at least the following elements:

- Audit of existing metadata: the goal is to review a representative sample of records to identify inconsistencies, formatting errors, empty or misused fields.
- Identification of metadata schemas in use: it is essential to identify both the base metadata schema and the various mappings and export capabilities to other schemas.

- Review of vocabularies and authority files: evaluate and validate the implementation of controlled vocabularies, such as thesauri, authority files, identifiers, and controlled lists, for item types, language fields, and date formats.

Once the assessment is complete, four fundamental processes are recommended to optimize repository metadata, which are described in the following section.

## 2.1. Data cleaning and refinement

This stage is essential for optimizing the visibility of content stored in a repository. High-quality metadata “should allow digital users to intuitively conduct the tasks such as identifying, describing, managing and searching data” (Ma et al., 2009, p. 1).

In this regard, data cleaning is a process that detects and corrects errors, inconsistencies, and incomplete fields, aiming to improve interoperability and user experience (Van-Hooland & Verborgh, 2015; Westbrook et al., 2012). This enhances accuracy and improves information retrieval through a system’s search tools.

Among the various tools available to improve metadata quality in repositories, OpenRefine (<https://openrefine.org>) stands out. It is a powerful open-source tool for cleaning, transforming, and reconciling messy tabular data. OpenRefine is particularly useful for standardizing and enriching metadata in libraries, archives, and research datasets.

To use OpenRefine in the data cleaning process, it is essential to: 1) export metadata in a format compatible with OpenRefine, such as CSV, TSV, Excel, or JSON; and 2) ensure that each row represents a resource (such as a document or image) and each column corresponds to an element of the metadata schema. For example, using Dublin Core (DC), the columns might include *dc:title*, *dc:creator*, *dc:subject*, among others.

This approach enables batch cleaning of records with inconsistencies, typos, or formatting errors. Below is an example figure of DC records that could potentially be improved using tools like OpenRefine.

**Figure 1**

*Example of DC records with potential for optimization through data cleaning.*

dc:title	dc:creator	dc:subject	dc:date	dc:language
History of Science	Laura Gonzalez	Science; history	2020-06-10	eng
history of science	Laura gonzález	science; history	10/06/2020	en
Artificial Intelligence	María Gómez	Technology, AI; Machine Learning	2023	english
Big Data		Massive data; analytics	2021-03-14	en
Machine Learning	Carlos Ruiz	machine learning, artificial intelligence	2024-11-25	en
Cybersecurity	Ana Lopez	information security	2021/05/03	English

It is worth noting that this process can be carried out regardless of the base metadata schema used by the system. To systematize metadata into a format accepted by OpenRefine, auxiliary tools such as MarcEdit (<https://marcedit.reeset.net>) can be used if the base schema is MARC21.

Once the records are imported, OpenRefine can correct inconsistencies such as unnecessary spaces, inconsistent capitalization, duplicates, and incorrect formats (especially in dates) as well as normalize terms using controlled vocabularies. Cells can be split, similar values

grouped, and transformations applied using GREL expressions<sup>1</sup>. Additionally, OpenRefine allows for the detection of missing values, format validation, and data reconciliation with external sources. Finally, the cleaned data can be exported in the desired format, ready for reuse or reintegration into the repository.

## 2.2. Implementation of controlled vocabularies

To enhance metadata with the goal of optimizing visibility and interoperability with other systems, controlled vocabularies play a significant role (Chipangila et al., 2024). On the one hand, they enhance the search experience within the repository itself by standardizing vocabulary in key fields, such as authors, language, and keywords. On the other hand, they enhance integration with external systems by using identifiers, controlled vocabularies, or lists, and predefined formats in key fields for interoperability.

To incorporate controlled vocabularies into digital repositories, the first step is to identify the metadata fields that could benefit from normalization. For example: authors, dates, language, keywords or subjects, and resource type.

Subsequently, it is necessary to select appropriate vocabularies to control specific fields. Below are some options categorized accordingly:

- To control subjects or keywords: Some alternatives include the Library of Congress Subject Headings (LCSH), the UNESCO Thesaurus, or the United Nations Bibliographic Information System (UNBIS) Thesaurus. Once the vocabulary to be integrated is selected, the next step is to integrate the thesaurus in a format such as SKOS (Simple Knowledge Organization System) or RDF (Resource Description Framework) into the platform. This integration will depend strictly on the platform and must ensure that subject or descriptor fields are mapped to the thesaurus terms (for example, the dc.subject field in Dublin Core or field 650 in MARC21). Finally, it is essential to verify that terms autocomplete correctly and are linked to their respective URLs.
- To control authorities, controlled vocabularies such as VIAF (Virtual International Authority File) or persistent identifiers like ORCID or ROR (Research Organization Registry) can be utilized. The integration process for these elements also depends strictly on the platform. Likewise, it must be ensured that the metadata schema fields related to authorities are correctly mapped to the vocabulary or list being integrated.
- To control item type: Various vocabularies exist to standardize item types and enhance interoperability, for example, with reference management systems. Notable examples include the DCMI Type Vocabulary (Dublin Core), COAR Resource Type Vocabulary, MODS Resource Types, Schema.org / CreativeWork Types, and OpenAIRE Guidelines Types, among others. After selecting the vocabulary, existing local values must be mapped to the chosen vocabulary, data entry forms should be adjusted to use controlled lists, and consistency of types across records must be ensured. Additionally, metadata export (e.g., in Dublin Core or XML) should be adapted to comply with interoperability standards and facilitate harvesting by aggregators. The process may include data validation, user interface adjustments, and testing to ensure that values are accurately reflected in both the internal administration interface (backend) and the public view of the repository.

---

1 GREL is a programming language designed to facilitate the organization, transformation, and querying of data in OpenRefine.



- To control key fields: normalizing key fields, such as language or publication date, is essential for standardizing metadata. Proper normalization improves the search experience within the repository and ensures effective data exchange with external services. For these fields, it is recommended to adopt ISO 8601 for date normalization and ISO 639 for language codes. This involves using the YYYY-MM-DD format for dates and two- or three-letter codes for languages. Implementation is achieved through controlled lists, automatic validation in data entry forms, and adjustments to the system's metadata templates, such as in Dublin Core, MARC21, or MODS (Metadata Object Description Schema). This ensures consistency, interoperability, and compatibility with external harvesters.

## 2.3. Periodic audits and adjustments according to regulatory changes

To carry out periodic metadata quality audits, it is recommended to establish a regular routine, such as quarterly or semiannually, in which representative samples of records are reviewed. During this review, consistency, completeness, correct use of controlled vocabularies, standardized formats (such as ISO 8601 and ISO 639-1) and the absence of typographical errors or duplicates should be verified. The use of automatic validation tools and quality report generation helps efficiently detect issues.

In response to changes in regulations or standards, it is essential to actively monitor updates in schemas such as Dublin Core, COAR, or OpenAIRE. When a modification occurs, the metadata mapping in the repository should be reviewed, and adjustments made to forms, vocabularies, or export templates. Staff should also be trained on the changes. Documenting each adjustment ensures traceability and facilitates future audits.

## 3. Enabling interoperability protocols

Enabling interoperability protocols in a digital repository is essential for increasing its visibility and reach. These protocols allow external services, such as aggregators, academic search engines, and national or regional portals, to automatically harvest the repository's metadata and content without manual intervention (Eells et al., 2024).

Thanks to this interoperability, the repository's records can be integrated into platforms such as OpenAIRE, BASE (Bielefeld Academic Search Engine), CORE (COncnecting REpositories), WorldCat, or even national or regional repositories, enhancing their discoverability for researchers, students, and the public. Additionally, it facilitates integration with other institutional systems and ensures that the content complies with open standards and open access policies.

There are several ways to promote interoperability through the following protocols:

- OAI-PMH Protocol: This protocol, based on HTTP and XML, enables other systems to harvest metadata from a repository automatically. It is widely used by aggregators such as OpenAIRE or BASE to collect information about available resources.
- REST API: This is an interface that allows other systems to query, create, or modify resources in a repository using HTTP requests, such as GET, POST, PUT, or DELETE. It is highly flexible and commonly used for integrations with external systems or custom applications.

- SWORD: This is a protocol that enables the remote deposit of content, such as articles, datasets, or theses, into a repository. It facilitates integration between publishing platforms, institutional systems, and digital repositories.

The activation of various protocols for interoperating with other systems strictly depends on the software and version being used. The available tools in each repository should be utilized to enable the API or SWORD, allowing for uploading or updating from external clients.

Likewise, each repository allows for the configuration of its OAI-PMH protocol, where the base URL of the service is defined and metadata is exposed in an appropriate format. In this context, metadata is essential and must be exposed correctly in formats readable by the protocol, such as Dublin Core, MARC, or DataCite.

The following table explains how to enable the OAI-PMH protocol according to each software platform:

**Table 2**  
OAI Protocol activation by software.

Software	How to Enable OAI-PMH?
DSpace <a href="https://wiki.lyrasis.org/display/DSDOC8x/OAI">https://wiki.lyrasis.org/display/DSDOC8x/OAI</a>	Although OAI-PMH is enabled by default, it is necessary to verify that "oai.enabled=true" and "oai.path=oai" are set in local.cfg or dspace.cfg.
EPrints <a href="https://wiki.eprints.org/w/OAI">https://wiki.eprints.org/w/OAI</a>	OAI support is enabled by default. You need to configure the base URL (oai.base_url), archive_id, sets, and any XSL stylesheets in cfg/cfg.d. Pay special attention to the oai.pl file.
Fedora <a href="https://github.com/saw-leipzig/foaipmh">https://github.com/saw-leipzig/foaipmh</a>	Fedora 6 does not include native OAI-PMH support. It requires implementing an external endpoint (e.g., Django + foaipmh) connected to its REST API.
InvenioRDM <a href="https://inveniordm.docs.cern.ch/reference/oai_pmh/">https://inveniordm.docs.cern.ch/reference/oai_pmh/</a>	OAI-PMH is enabled by default at /oai2d. From the admin interface, you can define sets and formats (oai_dc, oai_datacite).
TIND	From the admin panel, go to 'OAI Repository Admin', enable the provider, and define sets.
Digital Commons <a href="https://digitalcommons.elsevier.com/integration-preservation/digital-commons-and-oai-pmh">https://digitalcommons.elsevier.com/integration-preservation/digital-commons-and-oai-pmh</a>	OAI support is enabled by default. The exposed fields are configured through the metadata manager.
Dataverse <a href="https://guides.dataverse.org/en/latest/admin/harvestserver.html">https://guides.dataverse.org/en/latest/admin/harvestserver.html</a>	From the 'Harvesting Server' section in the Dashboard, the OAI-PMH service is enabled, and sets are defined. The endpoint is usually /oai.

To ensure the proper functioning of the OAI-PMH protocol, there are tools known as validators that evaluate the protocol's operability. One of the most well-known tools is the OAI-PMH Validator. This validator checks whether sets are adequately defined and whether records correctly export fields such as dates, identifiers, types, language, and other relevant information. It also allows for reviewing XML responses to detect errors or poorly structured formats. To analyze a repository's protocol functionality, simply enter the base URL, and the system will perform the analysis.

**Figure 2**

OAI-PMH Validator.

Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)  
Validator & data extractor Tool

Download and evaluate XML metadata from OAI-PMH enabled digital libraries.

Validate URL   Validate By Direct Input   Download XML   </> REST API   About   References

OAI-PMH URL:  Check now »  
Example OAI-PMH URL: https://oai.datacite.org/oai

AVAILABLE COMMANDS

- Identify
- ListMetadataFormats
- ListSets
- ListIdentifiers
- ListRecords MARCXML
- ListRecords OAI\_DC
- ListRecords OAI\_OPENAIRE

Identify Validation 7   XML Result 703 Bytes   Identify help

1. ☒ HTTP status 200
2. ☒ Content type text/xml
3. ☒ XML Content checked.
4. ☒ Request time is 0.298 sec
5. ☒ XML complies with OAI-PMH XML Schema <http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd>
6. ☒ OAI-PMH protocol version is 2.0.
7. ☒ Valid adminEmail library@wipo.int

To ensure the proper functioning of the protocol, it is necessary to carry out maintenance activities and periodically monitor its status. Additionally, it is essential to update metadata mappings when recommendations from COAR, OpenAIRE, or other aggregators are revised. For ongoing monitoring, it is also essential to document the endpoints<sup>2</sup> and maintain coordination with other systems that depend on the repository.

## 4. Adoption of persistent identifiers

A persistent identifier is a unique, durable, and resolvable digital reference to a specific object, such as an article, dataset, software, person, or organization. These identifiers are designed to remain valid and accessible over time, even if the object's physical location or hosting server changes (Meadows et al., 2019).

Typically, a persistent identifier has three essential components:

- Global uniqueness, which means it includes a controlled syntax and a namespace governed by clearly defined authorities;
- Persistence, which ensures stable links and resolution functions, as well as persistent schemas and referenced objects; and
- Resolvable for both humans and machines, providing information on how to find, access, or use the referenced object (De Castro et al., 2023).

Persistent identifiers (PIDs) are important for optimizing the visibility and citability of publications, as they make it easier for search engines, academic repositories, and analytics

<sup>2</sup> An endpoint is a specific address or URL through which an external system can interact with the repository to access its services or data.

tools to automatically find and link to documents without relying on unstable URLs. They also provide long-term stability by promoting resource accessibility, as they combat “link rot”<sup>3</sup> and “content drift”<sup>4</sup>: even if the object is moved, its persistent identifier will still resolve correctly. Finally, when PIDs are associated with structured metadata and supported by robust infrastructures, they enhance the trustworthiness and reputation of the content.

Among the most widely recognized PIDs is the Digital Object Identifier (DOI). It is the most used identifier for articles, books, datasets, and software. This system combines a permanent identifier with mandatory metadata and guaranteed resolution. When a DOI is resolved, it leads to a landing page with metadata, enhancing visibility and citation tracking. DOIs are assigned by registration agencies such as CrossRef and DataCite and are generally more costly than other identifiers.

Another option is the Handle system, a non-commercial identifier that has been used since 1995. Its main goal is to provide persistent identification and resolution services, operated centrally by the Corporation for National Research Initiatives (CNRI). A Handle identifier consists of a prefix that identifies the authority, along with a suffix that refers to the object being identified. Handle is the technical foundation of DOI and is more affordable to implement. Some systems, such as DSpace, integrate the Handle system by default; however, it must be acquired and configured to resolve resources through the identifier properly (<https://wiki.lyrasis.org/display/DSDOC8x/Handle.Net+Registry+Support>).

An alternative is the ARK (Archival Resource Key) persistent identifier, a system designed to provide durable and reliable links to digital objects, particularly useful in libraries, archives, and museums. Unlike other identifiers such as DOI or Handle, ARK is cheaper, decentralized, and highly flexible (<https://arks.org>), allowing institutions to generate and manage their identifiers without relying on a central registration authority. Its typical format is ark:/NAAN/identifier, where the NAAN identifies the issuing organization. A distinctive feature of ARK is its ability to provide access not only to the digital object but also to its metadata and a persistence commitment statement, which reinforces transparency and trust in long-term preservation. This system has been widely adopted by institutions such as the U.S. Library of Congress and the California Digital Library, supporting the visibility and traceability of cultural and academic resources.

**Table 3**  
*Comparison of persistent identifiers.*

	DOI	Handle	ARK
Management	Centralized (DataCite, Crossref)	Distributed (CNRI)	Decentralized (institutional)
Structure	10.1234/abc123	20.5000/xyz456	ark:/12345/x6789
Resolution	Yes, via <a href="https://doi.org/">https://doi.org/</a>	Yes, via <a href="https://hdl.handle.net/">https://hdl.handle.net/</a>	Yes, via <a href="https://n2t.net/">https://n2t.net/</a> or locally
Metadata Access	Yes (mandatory landing page)	Yes (depending on usage)	Yes (via inflection, i.e., a modifier character in the URL)
Guaranteed Persistence	High (by contract)	High (depends on repository)	Variable (based on institutional policy)

3 Link rot occurs when a hyperlink no longer leads to the intended content because the page has been moved, deleted, or the domain is no longer active.

4 Content drift occurs when the content at a given URL changes over time, so it no longer reflects what was originally cited or intended, even though the link still works.

	DOI	Handle	ARK
Cost	Requires paid membership and may include additional cost per DOI assignment	Annual fee of USD 50	No payment or membership required
Typical Uses	Articles, datasets, software	Repository objects (DSpace, Fedora)	Archives, libraries, digital museums

To implement a persistent identifier, the first step is to select which one will be used. It is worth noting that identifiers are not mutually exclusive and can be combined. For example, a resource can have both a DOI and an ARK.

It is necessary to register with a persistent identifier provider DOIs, this can be obtained through one of the registration agencies, such as Crossref or DataCite, although both require a paid membership and may involve a fee for minting each DOI.

The Handle system can be acquired through CNRI by paying USD 50 and linking it to a compatible repository system (<https://www.handle.net/payment.html>). Once configured, each deposited object receives an identifier with an authorized prefix, assigned by CNRI, and a unique suffix. These identifiers are resolved via <https://hdl.handle.net/>, ensuring long-term accessibility even if the resource's physical location changes.

ARK can be obtained by requesting a Name Assigning Authority Number (NAAN) at [arks.org](https://arks.org) (<https://arks.org/about/getting-started-implementing-arks>). A NAAN is a unique number that identifies the ARK-issuing institution within the system. It functions as an official prefix that ensures each organization creating ARK identifiers has its exclusive namespace.

Afterward, it is necessary to configure the repository to issue and maintain persistent identifiers. This will depend on the specific software used to manage the institutional repository and can be done through registration agencies, the use of plugins, integration within the system itself, or by managing key configuration files.

Finally, the assignment of PIDs must be integrated into the workflow for setting up new resources, specifically the "identifier" field. It is also essential to ensure that the identifier correctly resolves to a landing page for the object, including its metadata and access to the resource.

PIDs such as DOI, Handle, and ARK are fundamental tools for strengthening the visibility of documents in a digital repository. By providing stable, unique, and long-lasting links, they ensure that resources remain easily findable, accessible, and citable, even when their technical location changes over time. Moreover, by being integrated into global resolution infrastructures and associated with structured metadata, these identifiers facilitate discovery by search engines, harvesters, academic citation systems, and open data networks. Altogether, PIDs ensure that documents are not only preserved but also disseminated and recognized in today's digital environments.

## 5. SEO Optimization of the repository

Search Engine Optimization (SEO) in digital repositories is a key strategy for increasing the visibility, accessibility, and impact of the academic, scientific, and cultural content they host. Despite having structured metadata and preservation standards, many repositories fail to rank well in search engines like Google or Bing, limiting the organic discovery of their resources by

users.

Implementing SEO best practices, including proper use of HTML tags, exposing Dublin Core metadata in schema.org format, creating user-friendly URLs, generating sitemaps, and enabling automatic indexing, improves how search engines interpret content. These measures facilitate accurate understanding, categorization, and prioritization of documents.

Moreover, combining SEO with persistent identifiers, such as DOI, Handle or ARK, reinforces the stability and traceability of resources on the web.

In a digital environment where attention is limited and competition for visibility is high, optimizing a repository's SEO is not just a technical improvement, but a strategic action to ensure that deposited resources fulfill their mission of being found, used, and cited.

The following elements are recommended for optimizing SEO in a repository:

### 5.1. Proper use of semantic HTML tagging

Use semantic tags such as <title>, <meta name="description">, <meta name="citation\_doi">, <meta name="citation\_author">, <h1>, <h2>, <article>, <section>, among others, to help search engines understand the structure of the content. This also helps tools like Altmetric better track the metrics of a particular resource (<https://help.altmetric.com/support/solutions/articles/6000240582-required-metadata-for-content-tracking>) (Reyes-Lillo & Pastor-Ramon, 2024).

It is crucial to ensure that each resource page (such as a document) has a unique and descriptive <title>. Additionally, including enriched metadata using schema.org or Dublin Core embedded in <meta> tags or JSON-LD format is recommended.

Below, you can see an example of proper semantic tagging using Dublin Core:

**Figure 3**

*Example of Dublin Core embedded in <meta> tags.*

```
<head>
  <meta name="DC.title" content="How to improve your metadata" />
  <meta name="DC.creator" content="Laura Martinez" />
  <meta name="DC.date" content="2023-11-15" />
  <meta name="DC.identifier" content="http://doi.org/10.1234/test-doi.2025.001" />
  <meta name="DC.language" content="es" />
  <meta name="DC.type" content="Book" />
</head>
```

And an example of semantic tagging using JSON-LD:

**Figure 4**

*JSON-LD Example.*

```
<script type="application/ld+json">
{
  "@context": "https://schema.org",
  "@type": "Book",
  "name": "How to improve your metadata",
  "author": {
    "@type": "Person",
    "name": "Laura Martinez"
  },
  "datePublished": "2023-11-15",
  "identifier": {
    "@type": "PropertyValue",
    "propertyID": "Handle",
    "value": "http://doi.org/10.1234/test-doi.2025.001"
  },
  "inLanguage": "es",
  "license": "https://creativecommons.org/licenses/by/4.0/",
  "publisher": {
    "@type": "Organization",
    "name": "Universidad Nacional"
  }
}
</script>
```

## 5.2. Creating and maintaining an XML sitemap

Creating and maintaining an XML sitemap in a digital repository is essential for enhancing the indexing and visibility of content to search engines like Google or Bing. A sitemap acts as a structured map that lists all relevant pages of the repository, allowing search engines to discover new documents, updates, or deposited resources quickly.

For it to function correctly, the sitemap must include only public, permanent, and accessible URLs (such as those containing persistent identifiers like DOIs or Handles), and it should be updated automatically whenever content is added or modified. Additionally, it must be correctly referenced in the robots.txt file and submitted to tools like Google Search Console to maximize its effectiveness.

A well-implemented sitemap not only speeds up indexing but also improves the SEO ranking of resources, increasing their reach and usage within the academic and scientific ecosystem.

## 5.3. Proper configuration of the robots.txt file

The robots.txt file plays a key role in the SEO optimization of digital repositories, as it controls how search engines access, crawl, and index their contents. This file, located at the root of the website, tells robots (such as Googlebot or Bingbot) which parts of the repository can be explored and which should be excluded.

Proper configuration allows search engines to access the pages of digital objects, such as landing pages with persistent identifiers. On the other hand, it can block irrelevant or sensitive paths, such as administrative areas or navigation filters that could generate duplicate content. For this reason, the file must not block relevant paths linked to persistent identifiers, such as /handle/ or /ark:/.



A simple example is the following:

**Figure 5**

*Example of elements to consider in the robots.txt file.*

```
User-agent: *  
Allow: /handle/  
Disallow: /admin/  
Disallow: /private/  
Sitemap: https://repository.org/sitemap.xml
```

If not configured correctly, the robots.txt file can accidentally prevent the indexing of important resources, negatively affecting their visibility and discoverability in search results. Additionally, it should include a reference to the sitemap.xml file, making it easier for search engines to perform a structured crawl of the content.

#### 5.4. Apply SSR (Server-Side Rendering) or hybrid rendering

Applying SSR (Server-Side Rendering) or a hybrid rendering approach in a digital repository is a technical strategy that helps improve the site's visibility and performance, especially in an increasingly AI-driven and automated indexing web environment.

Unlike CSR (Client-Side Rendering), where content is dynamically generated in the browser, SSR allows pages to be generated on the server before being sent to the user or search engine bot. This has multiple benefits: first, it improves SEO, as search engines can immediately access structured content without relying on JavaScript to render it. Second, it speeds up initial load times, enhancing user experience and supporting navigation from mobile devices or slow networks.

Moreover, by implementing hybrid rendering, SSR combined with CSR, an optimal balance is achieved between performance, interactivity, and visibility, making the repository effective for both humans and indexing bots or generative engines.

There are tools like Next.js, Nuxt, or Rendertron that help adapt sites to be SEO-friendly and compatible with these rendering strategies.

### 6. Generative Engine Optimization in repositories: a factor to consider

Generative Engine Optimization (GEO) is an emerging concept that refers to the optimization of digital content for generative search engines, such as ChatGPT, Google Search Generative Experience (SGE), or Perplexity, which use artificial intelligence (AI) to respond with directly generated text, rather than simply displaying links as traditional SEO does (Daniels, 2025).

GEO aims to adapt the way digital content is structured and tagged so that language models (LLMs) can correctly interpret it, reference it in their responses, and integrate it into AI-generated content (Aggarwal et al., 2024).

Just as SEO optimizes content to be more visible on Google, GEO optimizes content to be understood, cited, and used by AI-powered answer generation engines.

In this context, repositories can also take specific actions to optimize their content based on GEO strategies. For example, structuring metadata using technologies like schema.org, JSON-LD, or Open Graph is highly recommended to support the inclusion of their content in AI-generated responses.

Additionally, it is essential to provide clear and accessible content, avoiding hidden or overly technical language that may be difficult for LLMs to comprehend. It is also necessary to clearly indicate flexible intellectual property licenses, such as Creative Commons, to facilitate content reuse.

Moreover, the use of descriptive landing pages, persistent identifiers to ensure traceability, providing RDF or JSON files, and enabling API access are key factors that help improve content for processing by LLMs.

In summary, GEO represents a new visibility paradigm for digital repositories, where it is no longer enough to appear on Google; content must be structured, accessible, and understandable by language models.. Implementing GEO strategies not only increases the reach of resources but also prepares the repository to integrate into the AI-based search and discovery ecosystem that is shaping the future of knowledge access.

## 7. Conclusion

Visibility optimization in digital repositories requires a comprehensive approach that combines technical, regulatory, and strategic aspects. The quality of metadata ensures that resources are understandable and reusable by both humans and machines; interoperability allows for their seamless integration into global information networks; and persistent identifiers guarantee their traceability and long-term access.

At the same time, strong SEO optimization improves search engine ranking, while incorporating a GEO perspective expands the reach of content to generative artificial intelligence agents.

Together, these elements strengthen the visibility, impact, and circulation of the knowledge hosted in repositories, aligning them with the principles of open science and equitable access to information.

## 8. Funding

This work is part of the Project “Parameters and strategies to increase the relevance of media and digital communication in society: curation, visualisation and visibility (CUVICOM)”. Grant PID2021-123579OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU.

## 9. References

Aggarwal, P., Murahari, V., Rajpurohit, T., Kalyan, A., Narasimhan, K., & Deshpande, A. (2024). *GEO: Generative Engine Optimization* (arXiv:2311.09735). arXiv. <https://doi.org/10.48550/arXiv.2311.09735>

- Chipangila, B., Liswaniso, E., Mawila, A., Mwanza, P., Nawila, D., M'sendo, R., Nyirenda, M., & Phiri, L. (2024). Controlled vocabularies in digital libraries: Challenges and solutions for increased discoverability of digital objects. *International Journal on Digital Libraries*, 25(2), 139–155. <https://doi.org/10.1007/s00799-023-00374-1>
- Daniels, C. (2025). *From SEO to GEO: How agencies are navigating LLM-driven search*. Campaign Asia. <https://www.campaignasia.com/article/from-seo-to-geo-how-agencies-are-navigating-llm-driven-search/501593>
- De Castro, P., Herb, U., Rothfritz, L., & Schöpfel, J. (2023). *Building the plane as we fly it: The promise of Persistent Identifiers*. Zenodo. <https://doi.org/10.5281/zenodo.7258286>
- Eells, L., Kelly, J., & Farrell, S. (2024). Repository (R)evolution: Metadata, interoperability, and sustainability. *Journal of Librarianship and Scholarly Communication*, 12(1), Article 1. <https://doi.org/10.31274/jlsc.16890>
- Ma, S., Lu, C., Lin, X., & Galloway, M. (2009). Evaluating the metadata quality of the IPL. *Proceedings of the American Society for Information Science and Technology*, 46(1), 1–17. <https://doi.org/10.1002/meet.2009.1450460249>
- Meadows, A., Haak, L. L., & Brown, J. (2019). Persistent identifiers: The building blocks of the research information infrastructure. *Insights*, 32(1). <https://doi.org/10.1629/uksg.457>
- OpenAIRE. (2020). OpenAIRE Guidelines—OpenAIRE Guidelines documentation. Retrieved from: <https://guidelines.openaire.eu/en/latest>
- Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J., & Haustein, S. (2018). The state of OA: A large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, 6, e4375. <https://doi.org/10.7717/peerj.4375>
- Reyes-Lillo, D., Morales-Vargas, A. & Rovira, C. (2025). Visibility, discoverability, findability, Search Engine Optimization (SEO) and Academic SEO in digital repositories: A scoping review. *BiD: textos universitaris de biblioteconomia i documentació*, (54). <https://doi.org/10.1344/BID2025.54.0>
- Reyes-Lillo, D., & Pastor-Ramon, E. (2024). Use of handle in institutional repositories and its relationship with alternative metrics: A case study in Spanish-speaking America. *Hipertext. Net*, 29, Article 29. <https://doi.org/10.31009/hipertext.net.2024.i29.17>
- Swan, A. (2010, February). *The open access citation advantage: Studies and results to date*. <https://eprints.soton.ac.uk/268516/>
- UNESCO. (2021). *UNESCO recommendation on open science*. <https://unesdoc.unesco.org/ark:/48223/pf0000379949>
- Van Hooland, S., & Verborgh, R. (2015). *Linked data for libraries, archives and museums: How to clean, link and publish your metadata*. Facet. <https://doi.org/10.29085/9781783300389>
- Westbrook, R. N., Johnson, D., Carter, K., & Lockwood, A. (2012). Metadata clean sweep: A digital library audit project. *D-Lib magazine*, 18(5/6). <https://doi.org/10.1045/may2012-westbrook>

# AI in Ibero-American Newsrooms: use cases and best practices

**Alexis Apablaza-Campos**

Universidad UNIACC, Chile

<https://orcid.org/0000-0003-4407-1145>

**Jaime Andrés Wilches Tinjacá**

Institución Universitario Politécnico Granacolombiano, Colombia

<https://orcid.org/0000-0002-4425-9394>

---

Apablaza-Campos, A., & Wilches Tinjacá, J. A. (2025). AI in Ibero-American Newsrooms: use cases and best practices. In J. Guallar, M. Vállez, & A. Ventura-Cisquella (Coords). *Digital communication. Trends and good practices* (pp. 134-147). Ediciones Profesionales de la Información. <https://doi.org/10.3145/cuvicom.10.eng>

---

## Abstract

This chapter examines the growing integration of artificial intelligence (AI) into journalistic routines across Ibero-America, accelerated by the emergence of generative models. Its aim is to explore the main uses of AI in content generation and to propose best practices. Seven key areas of application are identified: (1) virtual presenters, (2) automated newsrooms, (3) use of generative AI, (4) audio-to-text conversion, (5) assisted SEO tools, (6) article summarization, and (7) customized chatbots. Through numerous regional examples, it demonstrates that AI functions as a support tool that streamlines processes and enhances creativity, without replacing human labour. The conclusions highlight the importance of ethics, transparency, continuous training, adaptation to local contexts, and strategic collaborations for a balanced and responsible adoption of AI in journalism.

## Keywords

Artificial intelligence; Content generation; Generative artificial intelligence; News outlet; Ibero-America.

---

## 1. Introduction

In recent years, artificial intelligence (AI) has ceased to be a so-called “technology of the future” to become a central element in journalistic routines. The emergence of generative models such as ChatGPT, launched on November 30, 2022, marked a turning point that accelerated the adoption of AI tools in the news industry (Apablaza-Campos & Codina, 2023).

Recent studies show that news editors are using AI to automate content, improve distribution, create products under human supervision, and offer personalized recommendations to their audiences (Segarra-Saavedra et al., 2019; Apablaza-Campos, 2024; Newman & Cherubini, 2025). This rapid evolution presents opportunities, but also strategic and ethical challenges that news outlet managers and journalists must consider.

The aim of this chapter is to synthesize the main ways in which news outlets in Ibero-America are incorporating AI into content generation and, based on these uses, to propose best practices and recommendations for newsroom leaders, journalists, and individuals interested in the future of journalism.

This is a brief and practical text, focused on conclusions and suggestions rather than lengthy theoretical discussions. Its structure is based on seven major areas of application, identified mainly in the second edition of the book *Inteligencia artificial para la generación de contenidos: experiencias editoriales en medios de comunicación de Iberoamérica* (Apablaza-Campos & Wilches Tinjacá, 2025); each section describes the category, presents examples, and extracts useful lessons for the regional news industry.

## 2. Virtual presenters

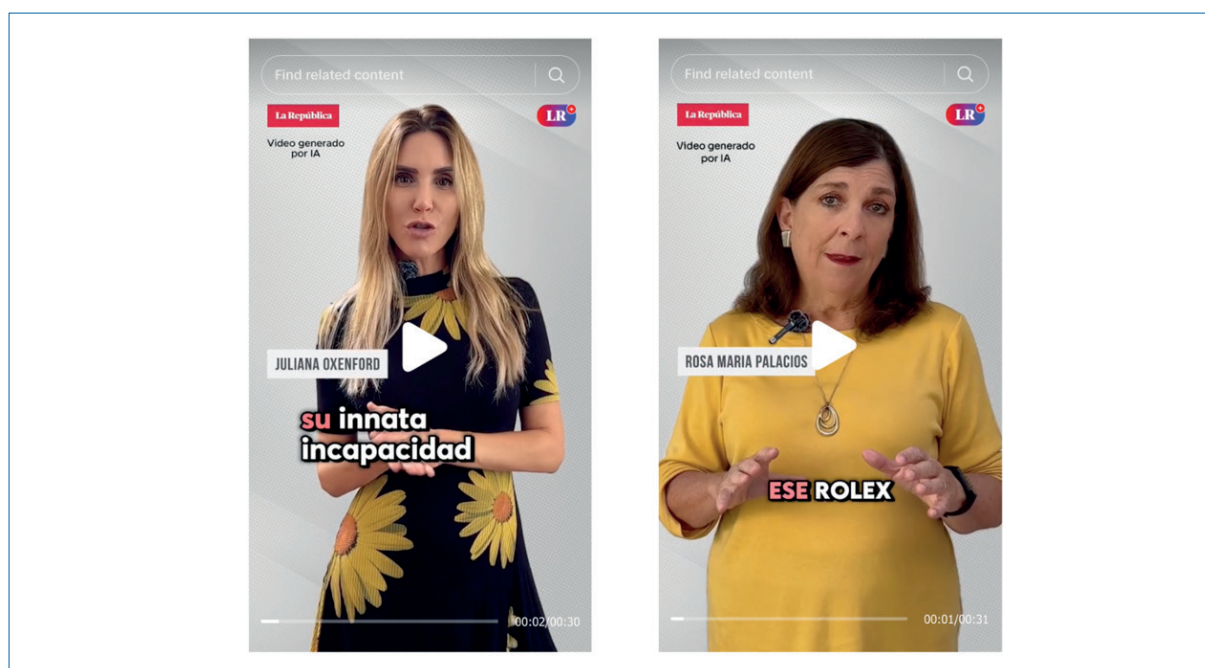
Virtual presenters or AI-generated avatars are synthetic models that “present” news in video or audio format. This technology, based on voice synthesis systems and facial animation, enables the creation of digital hosts capable of reading scripts, maintaining a consistent style, and being adapted to different languages. For news outlets, virtual presenters can help reduce audiovisual production costs, expand language offerings, and provide continuous event coverage.

### 2.1. Notable examples by country

- **México:** NAT, developed by *Grupo Fórmula*, became the country's first AI news anchor (Kusunoki, 2024). Another initiative is CLARA, an acronym for *Conductora Lógica de Asistencia y Respuesta Automática*, created by university media such as *Canal 44* and *Radio Universidad de Guadalajara* (Del Campo & Ramírez Santos, 2024).
- **Perú:** The university channel *Letras TV* launched *Illariy*, an avatar that presents newscasts in Quechua (Kusunoki, 2024). The newspaper *La República* uses avatars that replicate the voices of its columnists to turn articles into video columns published on TikTok (Lozano Chávez, 2024).
- **Other countries:** In Spain, *RTVE* uses *Hiperia*, a music content avatar (Salaverría, 2024); while television outlets such as *Megacadena* in Paraguay and *Radio Televisión Dominicana* are experimenting with avatars in their main newscasts or even to narrate historical events like the struggle for independence (Lora & Álvarez Álvarez, 2024; Ferreira Candia et al., 2025). In Ecuador, *Teleamazonas* allowed its audience to choose the name of an AI presenter for its technology segment (Espinosa, 2024).

**Figure 1**

AI-generated video columns by *La República* (Peru).



(Lozano Chávez, 2024).



## 2.2. Best practices and recommendations

- **Value linguistic and cultural diversity:** The examples of *Illariy* and *La República* show that virtual presenters can help bring news closer to Quechua-speaking communities and younger audience segments. Adapting avatars to local dialects or languages strengthens cultural identity and may enhance content relevance.
- **Transparency with the audience:** News outlets are advised to clearly indicate when synthetic presenters are being used and to explain their purpose. Transparency helps avoid perceptions of deception and sets appropriate expectations regarding the reliability of information.
- **Complement, don't replace:** Although avatars reduce production costs, it is essential that they coexist with human journalists. Human oversight ensures quality, provides editorial judgment, and helps prevent automated errors or biases (Ventura Pocino, 2021).

## 3. AI-powered automated newsrooms

AI-powered automated newsrooms use algorithms to autonomously generate texts from structured data. They are commonly applied in fields such as sports, finance, or election results, where formats tend to be repetitive and data is abundant (Segarra et al., 2019). These tools enable the rapid production of high volumes of news items, allowing journalists to focus on higher-value tasks (Apablaza-Campos, 2024).

### 3.1. Notable examples by application

- **Sports coverage:** The company DataFactory provides automatic writing services to news outlets in Chile (*El Rancagüino*), Peru (*Radio RPP*), Venezuela (*Meridiano*), and Colombia (*Noticias RCN*, *Gol Caracol*, *Revista Semana*, and *Caracol Radio*), generating real-time match reports for football games in various leagues (Apablaza-Campos & Wilches Tinjacá, 2025).
- **Election coverage:** In Brazil, *TV Globo* and *G1* used AI to narrate the results of the 2022 presidential elections (Barbosa & Costa Pinto, 2025); in Spain, *RTVE* implemented automated systems to report on local elections in towns with fewer than one thousand inhabitants (Apablaza-Campos, 2025).
- **Current affairs and technology:** News outlets such as *ABC Color* (Paraguay), *El Comercio and Gadgeros* (Ecuador) use automation for current news and articles on technological innovation (Ferreira Candia et al., 2025; Espinosa, 2024).

### 3.2. Best practices and recommendations

- **Prioritize data verification:** Algorithms rely on numerical sources; therefore, news outlets must ensure that data sources are reliable and free of errors. A system of human review or automated validation will prevent the publication of incorrect information.
- **Editorial voice customization:** Even when writing is automated, outlets can tailor the style, tone, and structure of texts to preserve the brand's identity. This reinforces coherence and fosters reader trust.



- **Strategic use:** Automation should be employed to free up human resources, allowing journalists to focus on reporting, analysis, and investigative pieces. Integrating AI into daily routines can boost productivity without compromising quality (Segarra-Saavedra et al., 2019; Murcia Verdú et al., 2022; Apablaza-Campos, 2024).

**Figure 2**

*El Rancagüino (Chile) offers AI-generated sports content on its website and social media, clearly notifying readers.*



(Erlandsen, 2024).

## 4. Experiences with generative AI

Generative AI tools — such as ChatGPT, Gemini, and similar platforms — enable users to draft content, suggest script ideas, and create multimedia material. They are particularly useful for accelerating research, writing scripts, and exploring novel approaches. When used with editorial discretion, these tools can enhance journalistic creativity.

### 4.1. Notable examples by country

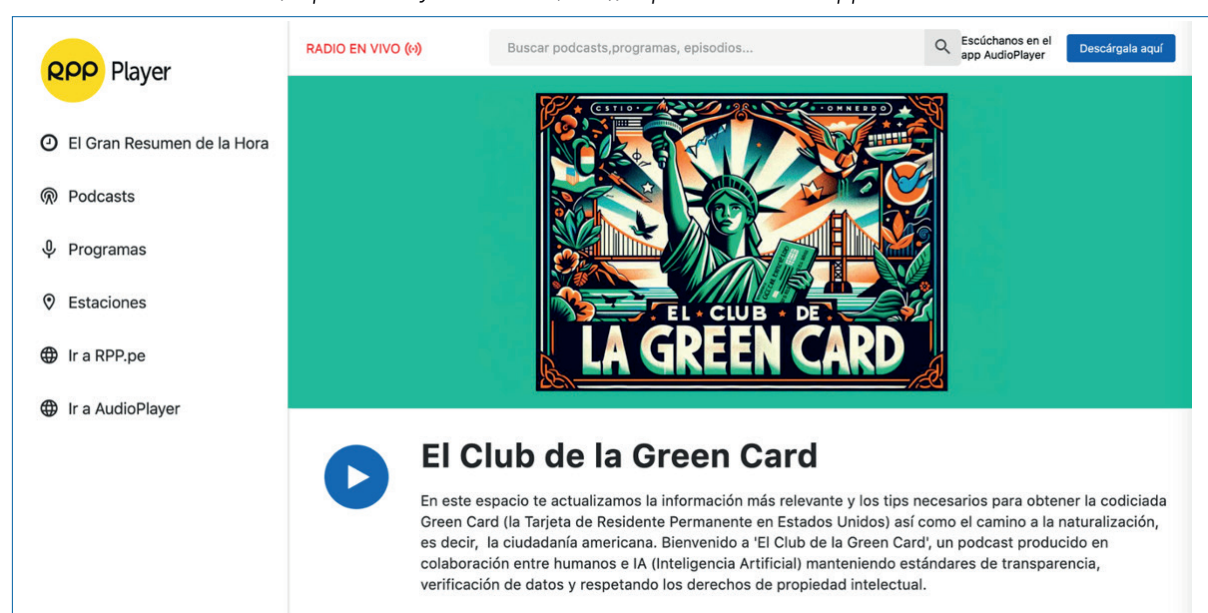
- **Bolivia:** The newspapers *Opinión* and *El Deber* use ChatGPT and Gemini for information gathering and research as part of the reporting process (Banegas Flores, 2024).
- **Guatemala:** The fact-checking outlet *Ojoconmipisto* uses AI to generate script ideas for its podcast (Alpírez, 2024).
- **Dominican Republic:** *Diario Libre* employs AI tools to optimize human-written content (Lora & Álvarez Álvarez, 2024).
- **Peru:** *RPP Radio* produces AI-assisted podcasts using ChatGPT (Lozano Chávez, 2024).
- **Uruguay:** *Canal 4* creates text and image content that is projected on screen (Roba, 2024).
- **Venezuela:** *2001online.com* uses various AI tools to generate videos that it shares on social media platforms (De Los Reyes & Sarmiento Altuve, 2025).

## 4.2. Best practices and recommendations

- **Use as a creative assistant:** Generative AI should not replace investigative work or human creativity, but rather, it should act as an assistant to suggest ideas, refine drafts, or translate texts. Journalists must validate AI-generated content and apply their professional judgment.
- **Bias and accuracy control:** Generative systems can produce biased or inaccurate responses. It is advisable to cross-check outputs with reliable sources and maintain a clear editorial protocol to determine which AI-generated inputs are integrated into the final product.
- **Training and literacy:** Teams should be trained in the use of these technologies, understand their limitations, and learn how to formulate effective prompts to obtain useful results (Gutiérrez Caneda et al., 2023).

**Figure 3**

*El Club de la Green Card, a podcast by RPP Radio (Peru), is produced with support from ChatGPT.*



(Lozano Chávez, 2024).

## 5. Audio conversion and voice synthesis

Audio conversion tools streamline the transcription of interviews or the generation of audio from text. They automate tedious tasks, enhance content accessibility, and open up new modes of consumption (such as automated podcasts or self-reading articles). These tools also include systems for translation and audio cleaning.

### 5.1. Notable examples by application

- **Interview transcription:** News outlets such as *El Deber* (Bolivia), *ABC Color* (Paraguay), *Canal 4* and *El Observador* (Uruguay) use AI to transcribe interview recordings (Apablaza-Campos & Wilches Tinjacá, 2025).

- **Texttospeech:** In Colombia, this system is widely used by outlets such as *El Espectador*, *El Tiempo*, *Blu Radio*, *Pulzo*, *El Heraldo de Barranquilla*, and *El Colombiano de Medellín* (Lozada Rodríguez, 2025).
- **Other applications:** *UOL* (Brazil) transcribes YouTube videos; *Infobae México* uses AI to clean and transcribe audio; *Quispe Chequea*, a project by *Ojo Público* in Peru, translates audio into Indigenous languages such as Quechua and Aymara; and Spanish media outlets like *Antena 3* and *Radio Bilbao* are experimenting with voice cloning (Apablaza-Campos & Wilches Tinjacá, 2025).

**Figure 4**

*El Espectador* (Colombia) integrates text-to-speech functionality into its content.



(Lozada Rodríguez, 2024).

## 5.2. Best practices and recommendations

- **Confidentiality and data protection:** When transcribing interviews, security protocols must be implemented to protect source identities and prevent leaks. Choosing tools that comply with privacy standards is essential.
- **Manual correction and editing:** AI-generated transcripts may contain errors; therefore, it is advisable to review and edit them to ensure their accuracy.
- **Adaptation to local languages:** Projects like *Quispe Chequea* highlight the importance of translating and synthesizing audio in Indigenous languages. This practice improves inclusivity and can inspire similar initiatives in other regions (Lozano Chávez, 2024).

## 6. Assisted SEO techniques

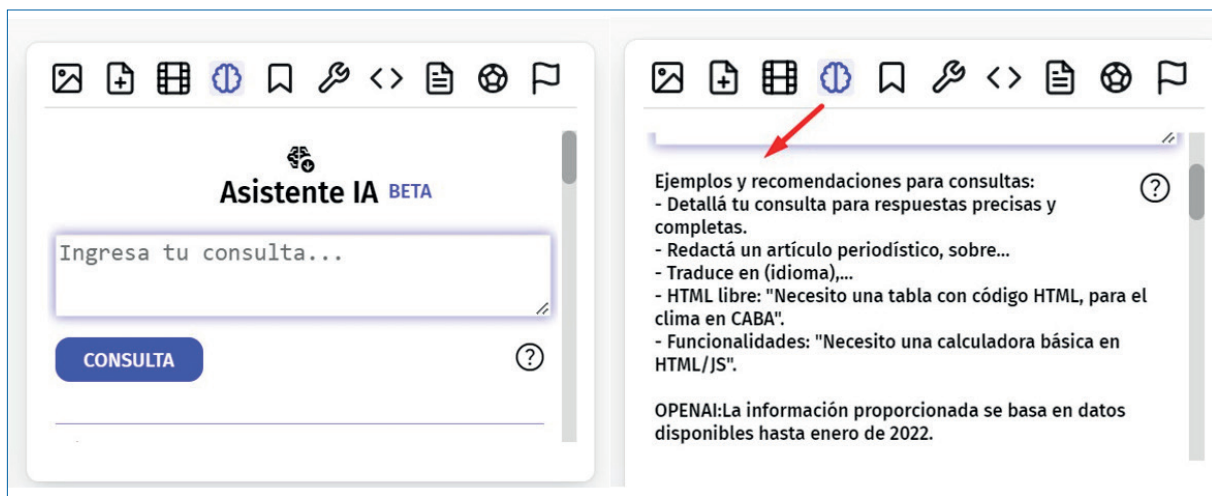
Assisted SEO refers to tools integrated into content management systems (CMS) that offer writing and optimization suggestions to improve article visibility in search engines. These tools help refine headlines, descriptions, and tags, and rely on AI to suggest keywords and structures.

## 6.1. Notable examples by application

- **CMS integration:** News outlets such as *Guatemala.com*, *Infobae México*, *ABC Color* (Paraguay), and *Meridiano* (Venezuela) have been pioneers in implementing these tools to facilitate the writing of digital journalism (Apablaza-Campos & Wilches Tinjacá, 2025).
- **Assisted SEO and generative AI:** In Argentina, outlets like *Olé* and *Clarín* use AI-based assistants that provide writing suggestions and recommendations to improve search engine positioning (Macías, 2024).

**Figure 5**

AI assistant offering SEO recommendations and access to tools such as ChatGPT, available to editors at *Olé* and *Clarín* (Argentina).



(Macías, 2024).

## 6.2. Best practices and recommendations

- **Integrate SEO into the editorial workflow:** Using these tools from the early stages of writing streamlines production and avoids the need for later adjustments.
- **Avoid formulaic automation:** While assisted SEO suggests structures and keywords, it is advisable to adapt these recommendations to the context and editorial voice to maintain narrative quality and avoid generic content.
- **Ongoing training:** Keeping teams up to date on SEO practices and understanding how they interact with search algorithms helps maximize both reach and content relevance (Apablaza-Campos et al., 2025).

## 7. Article summarization

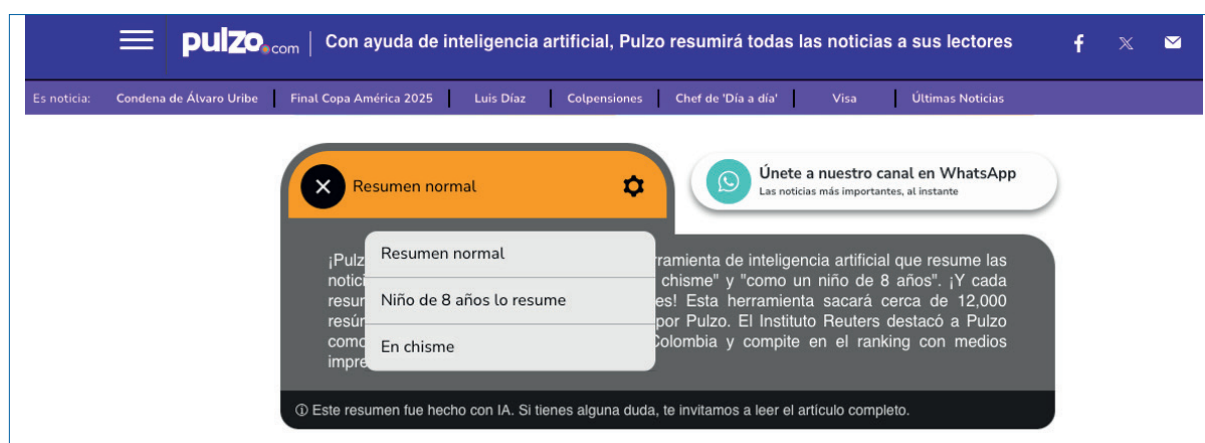
Automatic summarization tools generate condensed versions of long-form articles. Some offer varying levels of compression or adapt the language for specific audiences, allowing readers to choose how they consume information. These tools are useful for engaging audiences with limited time and making content more accessible to children or individuals with lower levels of media literacy.

## 7.1 Notable examples by application

- **Summarization modes:** While *BioBio Chile* condenses its longer content into a single paragraph (Erlandsen, 2024), *Pulzo* (Colombia) uses tools that summarize articles in three formats: standard, gossip-style, and a version for 8-year-olds (Lozada Rodríguez, 2025).
- **The UALTER Project:** An initiative by *Grupo Clarín* (Argentina) that offers multiple versions of the same text: a standard version, an abridged one with basic tips, and adaptations for younger readers or with colloquial language (Macías, 2024).

**Figura 6**

*Pulzo* (Colombia) offers three summary formats for its articles.



(Lozada Rodríguez, 2025).

## 7.2. Best practices and recommendations

- **Diversify formats:** Offering different summary levels (e.g., quick read, child-friendly version, or “gossip mode”) expands reach and allows content to be tailored to various audience segments.
- **Maintain accuracy:** Summarized versions must preserve key facts and context. Human oversight is necessary to ensure that the integrity of the news is not compromised.
- **Present clear options:** It is advisable to explicitly label content as a summary and provide a link to the full article, so readers can explore further if they wish.

## 8. Customized chatbots

AI-based conversational tools allow users to interact with a news outlet's content through natural language queries. These systems can function as fact-checking tools, search assistants, or thematic guides. Chatbots open new pathways for content distribution and foster a closer relationship between audiences and the media.

### 8.1 Notable examples by country

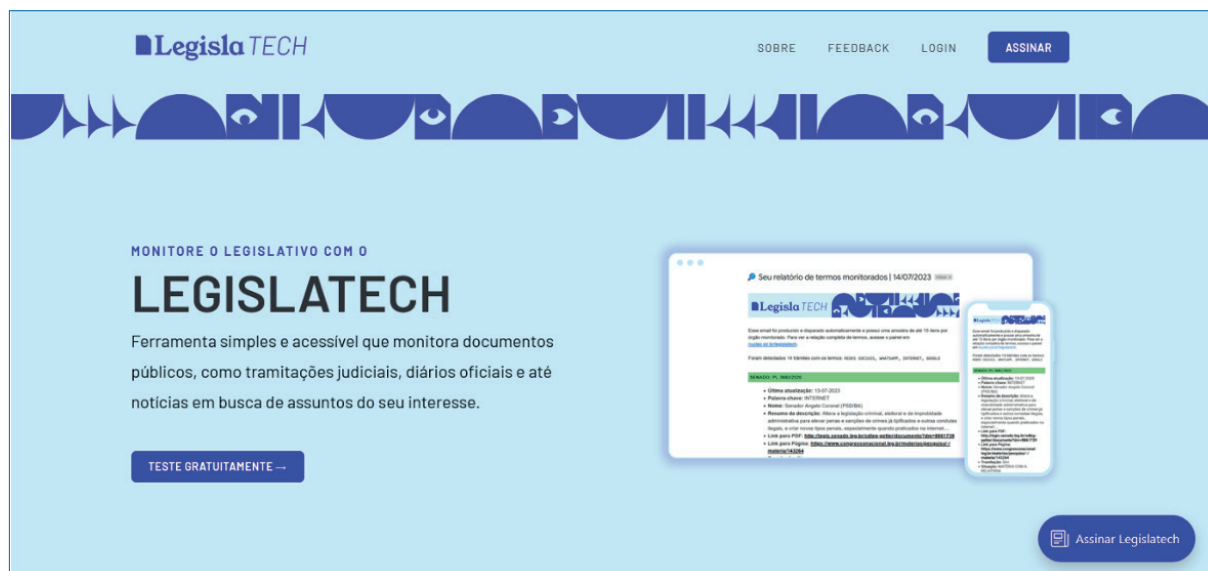
- **Brazil:** *FátimaGPT*, by Aos Fatos, and *Legislatech*, by Núcleo Jornalismo, analyze public documents and allow users to make personalized queries (Barbosa & Costa Pinto, 2025).



- Peru: *Quispe Chequea*, by *Ojo Público* — already mentioned in section 4 — uses customized GPTs to generate journalistic fact-checking texts (Lozano Chávez, 2024).

**Figure 7**

*Legislatch is a chatbot by Núcleo Jornalismo (Brazil), integrated into ChatGPT.*



(Barbosa & Costa Pinto, 2025)

## 8.2. Best practices and recommendations

- **Define the chatbot's purpose:** Some chatbots focus on providing access to public databases, others serve as fact-checking tools or as news consumption assistants. Before implementation, it is advisable to clearly define the specific need the chatbot is intended to address.
- **Ongoing updates and training:** Chatbots must integrate up-to-date information and adjust their responses based on user queries. Regular evaluations of accuracy and corrections for bias or errors are recommended.
- **Privacy and security:** Since users may share sensitive data, it is essential to comply with data protection regulations and communicate clearly how conversations are managed (Ventura Pocino, 2021).

## 9. General conclusions and future challenges

The review of use cases presented here demonstrates that artificial intelligence is already part of the Ibero-American media ecosystem (Infobae, 2024). Although scholarly research on the relationship between journalism and AI in the region remains incipient — with Spain accounting for 93% of indexed publications in databases such as Scopus or Web of Science (Apablaza-Campos et al., 2024) — the examples compiled here show a gradual adoption across Latin America, which has also been well documented in other works (Salas et al., 2023; Soto-Sanfiel et al., 2022; Navarro Zamora, 2023; Pinto & Barbosa, 2024).

AI appears in a wide variety of roles: content production, voice generation, translation, visibility optimization, and user service (Newman & Cherubini, 2025). This diversity of applications reveals that there is no single correct way to incorporate AI; each news outlet must evaluate its own resources, objectives, and audience (Apablaza-Campos & Wilches Tinjacá, 2025a).

From a strategic perspective, several cross-cutting ideas are worth highlighting::

- **AI as a support tool:** In all analyzed areas, AI complements human work. Automated newsrooms and virtual presenters speed up processes, but editorial oversight and creativity remain essential. AI should be understood as a tool to expand capabilities, not as a replacement for journalists (Apablaza-Campos & Codina, 2023; Apablaza-Campos, 2024).
- **Ethics and transparency:** Public trust depends on news outlets being transparent about their use of AI, informing users of its limitations, and upholding a commitment to truthfulness. It is crucial to establish clear policies regarding content authorship and personal data protection (Ventura Pocino, 2021).
- **Continuous training:** The rapid evolution of these technologies calls for training programs for journalists and media managers (Gonçalves & Melo, 2022). AI literacy will contribute to more responsible usage and help identify new opportunities.
- **Adaptation to local contexts:** The success of projects like *Illariy* or *Quispe Chequea* lies in their attention to local languages and realities. News outlets must consider their communities' needs and leverage AI to strengthen their connection with audiences (Kusunoki, 2024; Lozano Chávez, 2024).
- **Strategic collaborations:** Many successful experiences (such as DataFactory or partnerships with AI platforms) arise from alliances between news outlets and technology companies. Such collaborations can facilitate access to advanced capabilities and reduce costs (Apablaza-Campos & Wilches Tinjacá, 2025).

In the near future, news outlets in Ibero-America will face the challenge of incorporating increasingly sophisticated tools without undermining journalistic labour or fueling fears of human replacement. The development of public policies, ethical codes, and regulatory frameworks will be key to fostering a balanced adoption of AI (Ventura Pocino, 2024). The combination of technological innovation and social commitment will allow journalism to seize the opportunities of AI while also contributing to the fight against disinformation.

## 9. Funding

This work is part of the Project "Parameters and strategies to increase the relevance of media and digital communication in society: curation, visualisation and visibility (CUVICOM)". Grant PID2021-123579OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU.

## 10. References

Alpírez, A. C. (2024). Guatemala: IA: más tiempo para más periodismo. In A. Apablaza-Campos, & J. A. Wilches Tinjacá (Eds.), *Inteligencia artificial para la generación de contenidos en Iberoamérica: experiencias editoriales en medios de comunicación* (pp. 97–99). DataFactory, Institución Universitaria Politécnico Grancolombiano & Iniciación Científica.



Apablaza-Campos, A. (2024). Redacción humana versus redacción con inteligencia artificial automatizada: el caso de diario El Rancagüino y las notas algorítmicas. *Hipertext.net*, (28), 143-152. <https://doi.org/10.31009/hipertext.net.2024.i28.11>

Apablaza-Campos, A. (2025). Intel·ligència artificial en mitjans de comunicació: l'experiència de RTVE a les eleccions al Parlament de Catalunya del 2024. *Comunicació. Revista de Recerca i d'Anàlisi*, 42(1), 9–29. <https://doi.org/10.2436/20.3008.01.250>

Apablaza-Campos, A., & Codina, L. (2023). *ChatGPT en medios digitales: experiencias periodísticas con inteligencia artificial generativa*. Barcelona: Departament de Comunicació, Universitat Pompeu Fabra. Communication Reports 07. <https://doi.org/10.31009/cr.2023.07>

Apablaza-Campos, A.; Codina, L.; & Lopezosa, C. (2025). SEO de contenidos e inteligencia artificial: experiencias en medios digitales. In E. Ortega Fernández, M. J. Ufarte Ruiz, & La Rosa Barrolleta, L. (Eds.), *La integración de la inteligencia artificial en la comunicación digital* (pp. 335-350). Editorial Tirant Humanidades. <http://eprints.rclis.org/47034/>

Apablaza-Campos, A., & Wilches Tinjacá, J.A. (Eds.) (2025). *Inteligencia artificial para la generación de contenidos en Iberoamérica: experiencias editoriales en medios de comunicación* (1ª ed. Impresa & 2ª ed. Digital). DataFactory, Institución Universitaria Politécnico Grancolombiano & Iniciación Científica. <https://doi.org/10.15765/librosic.v5i60>

Apablaza-Campos, A., Wilches Tinjacá, J. A. & Salaverría, R. (2024). Generative artificial intelligence for journalistic content in IberoAmerica: perceptions, challenges and regional projections. *BiD*, 52. <https://doi.org/10.1344/BID2024.52.06>

Banegas Flores, N. C. (2024). Bolivia: entusiasmo y escepticismo ante la IA. In A. Apablaza Campos, & J. A. Wilches Tinjacá (Eds.), *Inteligencia artificial para la generación de contenidos en Iberoamérica: experiencias editoriales en medios de comunicación* (pp. 72–74). DataFactory, Institución Universitaria Politécnico Grancolombiano & Iniciación Científica.

Barbosa, S. & Costa Pinto, M. (2025). Brasil: entre ejemplos significativos y dificultades tecnológicas y financieras (trad. de J. Schena). In A. ApablazaCampos, & J. A. Wilches Tinjacá (Eds.), *Inteligencia artificial para la generación de contenidos en Iberoamérica: experiencias editoriales en medios de comunicación* (pp. 81–85). DataFactory, Institución Universitaria Politécnico Grancolombiano & Iniciación Científica.

De Los Reyes, L. & Sarmiento Altuve, Y. J. (2025). Venezuela: primeros (e incipientes) pasos periodísticos. In A. Apablaza-Campos, & J. A. Wilches Tinjacá (Eds.), *Inteligencia artificial para la generación de contenidos en Iberoamérica: experiencias editoriales en medios de comunicación* (pp. 120–123). DataFactory, Institución Universitaria Politécnico Grancolombiano & Iniciación Científica.

Del Campo, A.M. & Ramírez Santos, F. (2024). México: mucha cobertura, poca integración y nada de desarrollo. In A. Apablaza-Campos, & J. A. Wilches Tinjacá (Eds.), *Inteligencia artificial para la generación de contenidos en Iberoamérica: experiencias editoriales en medios de comunicación* (pp. 100–103). DataFactory, Institución Universitaria Politécnico Grancolombiano & Iniciación Científica.

Erlandsen, P. (2024). Chile: pasos tímidos y en secreto. In A. Apablaza-Campos, & J. A. Wilches Tinjacá (Eds.), *Inteligencia artificial para la generación de contenidos en Iberoamérica: experiencias editoriales en medios de comunicación* (pp. 83–85). DataFactory, Institución Universitaria Politécnico Grancolombiano & Iniciación Científica.

Espinosa, C. (2024). Ecuador: desde la automatización hacia la transformación de contenidos. In A. Apablaza-Campos, & J. A. Wilches Tinjacá (Eds.), *Inteligencia artificial para la generación de contenidos en Iberoamérica: experiencias editoriales en medios de comunicación* (pp. 90–93). DataFactory, Institución Universitaria Politécnico Grancolombiano & Iniciación Científica.

Ferreira-Candia, R., Daporta, N., Jara, R., & Sales, C. (2025). Paraguay: experiencias a paso lento. In A. Apablaza-Campos, & J. A. Wilches Tinjacá (Eds.), *Inteligencia artificial para la generación de contenidos en Iberoamérica: experiencias editoriales en medios de comunicación* (pp. 107–109). DataFactory, Institución Universitaria Politécnico Grancolombiano & Iniciación Científica.

Gonçalves, A. & Melo, P. (2022). Artificial Intelligence and Journalism: an Approach to the Portuguese Context. *Fonseca*, 25, 23–34. <https://doi.org/10.14201/fjc.29682>

Gutiérrez-Caneda, B., Vázquez-Herrero, J., & López-García, X. (2023). AI application in journalism: ChatGPT and the uses and risks of an emergent technology. *Profesional de la información*, 32(5). <https://doi.org/10.3145/epi.2023.sep.14>

Infobae (24 de junio de 2024). Apablaza Campos (docente de la UNIACC): “las rutinas periodísticas han tenido un proceso de evolución con la IA”. <https://www.infobae.com/america/agencias/2024/06/24/apablaza-campos-docente-de-la-uniacc-las-rutinas-periodisticas-han-tenido-un-proceso-de-evolucion-con-la-ia>

Kusunoki, J. (2024). América Latina: generar valor, un desafío regional para la IA. In A. Apablaza-Campos, & J. A. Wilches Tinjacá (Eds.), *Inteligencia artificial para la generación de contenidos en Iberoamérica: experiencias editoriales en medios de comunicación* (pp. 66–68). DataFactory, Institución Universitaria Politécnico Grancolombiano & Iniciación Científica.

Lora, E., & Álvarez Álvarez, A. (2024). República Dominicana: ¿probar “lo prohibido” en la redacción? In A. Apablaza-Campos, & J. A. Wilches Tinjacá (Eds.), *Inteligencia artificial para la generación de contenidos en Iberoamérica: experiencias editoriales en medios de comunicación* (pp. 110–113). DataFactory, Institución Universitaria Politécnico Grancolombiano & Iniciación Científica.

Lozada Rodríguez, A. D. (2025). Colombia: manifiestos y (auto)contenido: un ‘Macondo’ generativo. In A. Apablaza-Campos, & J. A. Wilches Tinjacá (Eds.), *Inteligencia artificial para la generación de contenidos en Iberoamérica: experiencias editoriales en medios de comunicación* (pp. 89–92). DataFactory, Institución Universitaria Politécnico Grancolombiano & Iniciación Científica.

Lozano Chávez, F. (2024). Perú: primeros pasos para seguir avanzando. In A. Apablaza-Campos, & J. A. Wilches Tinjacá (Eds.), *Inteligencia artificial para la generación de contenidos en Iberoamérica: experiencias editoriales en medios de comunicación* (pp. 107–109). DataFactory, Institución Universitaria Politécnico Grancolombiano & Iniciación Científica.

Macías, D. (2024). Argentina: la necesidad de defender nuestra identidad. In A. Apablaza-Campos, & J. A. Wilches Tinjacá (Eds.), *Inteligencia artificial para la generación de contenidos en Iberoamérica: experiencias editoriales en medios de comunicación* (pp. 69–71). DataFactory, Institución Universitaria Politécnico Grancolombiano & Iniciación Científica.

Murcia Verdú, F., Ramos Antón, R., & Calvo Rubio, L. (2022). Análisis comparado de la calidad de crónicas deportivas elaboradas por inteligencia artificial y periodistas. *Revista Latina de Comunicación Social*, 80, 91–111. <https://doi.org/10.4185/RLCS-2022-1553>

- Navarro Zamora, L. (2023). Inteligencia Artificial en el Ciberperiodismo de América Latina: Estudio exploratorio de las prácticas de los cibermedios y del ciberperiodista. *Question/Cuestión*, 3(75), e806. <https://doi.org/10.24215/16696581e806>
- Newman, N., & Cherubini, F. (9 de enero de 2025). *Journalism, media, and technology trends and predictions 2025*. Reuters Institute for the Study of Journalism. <https://reutersinstitute.politics.ox.ac.uk/journalism-media-and-technology-trends-and-predictions-2025>
- Pinto, M. C., & Barbosa, S. O. (2024). Artificial Intelligence (AI) in Brazilian Digital Journalism: Historical Context and Innovative Processes. *Journalism and Media*, 5(1), 325–341. <https://doi.org/10.3390/journalmedia5010022>
- Roba, N. (2024). Uruguay: IA, aliada y sospechada. In A. Apablaza-Campos, & J. A. Wilches Tinjacá (Eds.), *Inteligencia artificial para la generación de contenidos en Iberoamérica: experiencias editoriales en medios de comunicación* (pp. 114–116). DataFactory, Institución Universitaria Politécnico Grancolombiano & Iniciación Científica.
- Salas, M., Siles, I., & Carazo, C. (2023). Imaginarios de periodistas costarricenses sobre la inteligencia artificial y otras tecnologías digitales. *Textual & Visual Media*, 17(1), 61–75. <https://doi.org/10.56418/txt.17.1.2023.4>
- Salaverría, R. (2024). España: mucho interés, poca innovación. In A. Apablaza-Campos, & J. A. Wilches Tinjacá (Eds.), *Inteligencia artificial para la generación de contenidos en Iberoamérica: experiencias editoriales en medios de comunicación* (pp. 97–99). DataFactory, Institución Universitaria Politécnico Grancolombiano & Iniciación Científica.
- Segarra-Saavedra, J., Cristófol, F., & Martínez-Sala, A. (2019). Inteligencia artificial aplicada a la documentación informativa y redacción periodística deportiva. El caso de BeSoccer. *Doxa Comunicación*, (29), 275–286. <https://doi.org/10.31921/doxacom.n29a14>
- Soto-Sanfiel, M., Ibiti, A., Machado, M., Marín-Ochoa, B., Mendoza, M., Rosell, C., & Angulo-Brunet, A. (2022). In search of the Global South: assessing attitudes of Latin American journalists to artificial intelligence in journalism. *Journalism studies*, 23(10), 1197–1224. <https://doi.org/10.1080/1461670X.2022.2075786>
- Ventura Pocino, P. (2021). *Algoritmos en las redacciones: Retos y recomendaciones para dotar a la inteligencia artificial de los valores éticos del periodismo*. Barcelona: Consell de la Informació de Catalunya
- Ventura Pocino, P. (2024). Lo más importante es dar sentido de uso a la IA poniéndola al servicio de la misión periodística In A. Apablaza-Campos, & J. A. Wilches Tinjacá (Eds.), *Inteligencia artificial para la generación de contenidos en Iberoamérica: experiencias editoriales en medios de comunicación* (pp. 29–35). DataFactory, Institución Universitaria Politécnico Grancolombiano & Iniciación Científica.

# AI and image banks: A research methodology

**Pere Freixa**

Universitat Pompeu Fabra, Spain

<https://orcid.org/0000-0002-9199-1270>

**Mar Redondo-Arolas**

Universitat de Barcelona, Spain

<https://orcid.org/0000-0002-0000-8593>

**Lluís Codina**

Universitat Pompeu Fabra, Spain

<https://orcid.org/0000-0001-7020-1631>

**Carlos Lopezosa**

Universitat de Barcelona, Spain

<https://orcid.org/0000-0001-8619-2194>

---

Freixa, P., Redondo-Arolas, M., Codina, L., & Lopezosa, C. (2025). AI and image banks: A research methodology. In J. Guallar, M. Vázquez, & A. Ventura-Cisquella (Coords). *Digital communication. Trends and good practices* (pp. 148-160). Ediciones Profesionales de la Información. <https://doi.org/10.3145/cuvicom.11.eng>

---

## Abstract

This chapter presents a methodological framework for analysing gender bias and the presence of sociocultural stereotypes in professional stock image banks, with a specific focus on the visual results returned by photographic and AI-generated platforms. The study is based on the hypothesis that neutral prompts — those lacking explicit references to gender, age, or ethnicity — should, in the absence of cultural or technical bias, yield a balanced visual representation across different social categories. Any significant deviation from such proportionality may indicate the existence of implicit biases or recurrent visual clichés. To explore this, the authors analysed images retrieved from four professional platforms — two based on conventional photography and two relying on AI image generation. A system of coded indicators was developed to classify the representations in terms of gender, age, ethnicity, functional diversity, beauty norms, and depicted actions. The methodology excluded group images and near-identical variants to ensure diversity and analytical rigour. The findings reveal that AI-based platforms more consistently align with user prompts (60.36%) compared to traditional photographic databases (44.84%). However, both types of platforms exhibit stereotypical patterns, suggesting a persistence of visual tropes and clichés. The proposed methodology proves effective in detecting these biases and offers a transferable analytical framework. The chapter aims to contribute to broader efforts towards more inclusive visual cultures, encouraging further interdisciplinary research on algorithmic image generation and representation in digital media.

## Keywords

Gender bias; Stereotypes; Stock image platforms; Artificial intelligence; Visual representation; Image prompts; Algorithmic interpretation; Iconographic analysis; Media representation.

---

## 1. Introduction

Image banks are among the main resources used by the media to visually complement the content they publish (Codina, 2011; Kamin, 2023). According to several authors (Gynnild, 2017; Mortensen & Gade, 2023), the downsizing of professional photography staff in traditional media outlets has led to a growing reliance on stock images to illustrate journalistic content, at the expense of commissioned photography (Mortensen et al., 2023; Ferry, 2023; Mortensen et al., 2024; Hugues, 2024). In his pioneering study, Tsang (1984) had already observed that by the late 1970s and early 1980s, only about one quarter of the photographs published were produced in-house. The remainder originated from agencies, image banks, and freelance photographers. It is therefore not surprising that image banks, alongside news agencies, have become the visual resource with the greatest capacity to influence and establish visual standards (Machin, 2004; Machin & Polzer, 2015; Frosh, 2015, 2020).

Companies specialising in the distribution of stock photography have significantly expanded their capacity to generate all kinds of illustrations and visual resources — both photorealistic and fictional — through the incorporation of artificial intelligence (AI) technologies (Perdices-Castillo & Perianes-Rodríguez, 2011; Codina & Lopezosa, 2020; Vrabič-Dežman, 2024). AI

has also enabled the emergence of new players in the visual information market, who use image-generation tools to produce vast amounts of visual content. Through search interfaces, media outlets can access, locate, and select the images that best suit their needs within a highly competitive market (Allard, 2023; Bright, 2023; Sojit-Pejcha & Crabapple, 2023; Freixa & Redondo-Arolas, 2023; 2024).

The growing prominence of stock images in the media has elicited mixed responses. Critics of the use of both stock images and AI-generated content in journalism argue, among other points, that such resources reproduce clichés and stereotypes, thus further perpetuating biases and undermining media credibility (Mortensen et al., 2023; Aiello et al., 2023). Moreover, they contend that these images lack editorial or documentary value, as they are not tied to specific events (Mortensen et al., 2024). As Frosh (2003) describes, “The most striking feature of stock images is their close relationship with classification. Inscribed within advertising ideology, these images participate in a representation of reality shaped by selective categorisation” (p. 91).

Nonetheless, while providers of stock imagery are often criticised for reproducing standardised and stereotypical codes and formulas — regardless of the source of the photographs — Kwak and An (2016) demonstrated, using deep learning systems and large datasets, how media outlets use published press images to convey specific messages and often reinforce clichés and biases, irrespective of the images’ origins.

Concerns about the reproduction of biases —be they gender, racial, cultural, or ideological— in press imagery have been widely acknowledged and explored both in general terms (Miller, 1975; Luebke, 1989; Rodgers & Thorson, 2000; Rodgers et al., 2007; Thurlow et al., 2020; Freixa et al., 2025) and in the context of political communication (Waldman & Devitt, 1998; Goodnow, 2010; Rönnback et al., 2025).

These studies have employed combined qualitative and quantitative analytical methodologies, including statistical counting and both iconographic and iconological content analysis. Often, image analysis has been complemented by the examination of the accompanying text. Miller’s pioneering work, for example, involves counting the number of times men and women appear in the *Los Angeles Times* and *The Washington Post* over the course of a year, identifying the newspaper sections in which they are published and the roles depicted (Miller, 1975). By contrast, Goodnow (2010) focuses his study on a smaller sample of images, subjecting them to a detailed semiotic analysis that delves into the reading and interpretative codes inherent in journalistic imagery. As noted, Kwak and An (2016) adopted a quantitative approach, working with a much larger sample comprising two million images. Thurlow et al. (2020) and Freixa et al. (2025), meanwhile, applied semiotic analysis to samples of 600 images to assess the presence of stereotypes on stock image platforms.

This chapter presents a methodological proposal for conducting visual research on images, clearly defining the observation parameters and the sample size. The proposed system has proven effective in studies examining gender bias and stereotypes in both photographic stock images and AI-generated visuals (Freixa et al., 2025).

## 2. Image banks and search platforms

Image banks rely on digital platforms to make their collections available to clients. Through search interfaces, users can access, locate, and select those images that best suit their needs. This is a highly competitive market. Searches (conducted either through keywords or textual



prompts) generate visual galleries in response, presented as sets of thumbnail images which users can then select and download. Depending on the complexity of the search, these systems may return anything from a handful to an overwhelming number of images.

To develop the proposed methodological model, we begin with the assumption that the pre-selections offered by these systems are designed to align with user queries. However, they may also reveal algorithmic biases and reflect the presence of stereotypes, such as those related to gender or race, by prioritising certain visual constructs over others.

The working premise is that these platforms deploy trained algorithms to provide users with the most relevant possible results — those that best match their expectations. With some variations, nearly all search interfaces offer functionalities to contextualise and refine searches. Some of these features are technical, such as image size and format; others are commercial, such as copyright-related options. In addition, many platforms allow users to specify preferences related to attributes such as age, ethnicity, or gender of the individuals represented. All of them permit the use of keywords to help obtain more precise results.

Let us consider the simplest —and most common— scenario: what happens when a search is conducted using an especially neutral prompt, without any indication of a specific group of people to be represented, and without using the tools available to filter or parameterise the search? How does the system operate to ensure that it meets the client's expectations?

The hypothesis is that the less information provided to the platform, the greater the degree of interpretative responsibility the system must assume to satisfy the query. When a prompt is vague or ambiguous, platforms tend to return numerous results with a wide range of nuances and variations, thereby increasing the likelihood that at least some of them will match the user's needs.

This methodology is specifically designed to provoke such a situation. We argue that this scenario provides the most suitable research environment in which to observe the presence of clichés, biases, and stereotypes within image banks. When a neutral prompt is submitted to a platform —without any additional parameters— the system is compelled to deliver a large and varied selection of responses to accommodate a broad range of potential users. The observation, classification, and characterisation of the variables produced by the system in response to an explicitly neutral request can reveal dominant visual constructs, highly stereotyped patterns of response, and ultimately expose the presence of biased codes and encoding systems.

### **3. Platform selection, prompt definition, and sample size**

The commercialisation of stock images takes place in an extremely competitive market, where a wide array of platforms offer their services to both private users and media organisations. These may be either paid or free of charge, and may specialise in either conventional photography or AI-generated imagery. Based on the criterion of professional reputation (Piironen, 2022; 2023; May, 2024), four professional platforms were selected for this research — two specialising in traditional photographic images and two offering AI-generated visuals. It was also ensured in the selection process that each platform is capable of returning at least 50 images in response to a given query, within the visual gallery provided, without requiring the user to reformulate the search or modify the prompt.



**Table 1**

*Selection of professional stock image platforms*

Photographic image banks		AI-Generated image banks	
Shutterstock	<a href="https://www.shutterstock.com/photos">https://www.shutterstock.com/photos</a>	Lexica (Stable diffusion)	<a href="https://lexica.art/">https://lexica.art/</a>
Getty Images	<a href="https://www.gettyimages.com/">https://www.gettyimages.com/</a>	Adobe Stock (opción IA)	<a href="https://stock.adobe.com/es/">https://stock.adobe.com/es/</a>

Source: Own elaboration.

To formulate the search queries, we opted to use the natural language prompt functionality offered by these platforms, rather than relying on search equations based on Boolean operators (Battelle, 2006; Abadal & Codina, 2008; Codina, 2018). As previously mentioned, our hypothesis is that the use of neutral prompts activates the algorithmic interpretation systems embedded in these applications, to deliver what the system deems to be the most appropriate response to the request.

This approach is considered particularly suitable for detecting potential biases and stereotypes: the results returned by the system are likely to align with what is considered common, normative, or standardised within its database or training corpus.

The decision to use natural language prompts instead of search equations is motivated by three main reasons. First, because of their potential to trigger the algorithmic interpretive logic we aim to study. Second, because we are interested in exploring and testing this mode of interaction with image banks, given its emerging and innovative nature. And third, because we wish to analyse what occurs when users, despite sometimes having the option to carry out Boolean searches, choose instead to interact through prompts.

To construct the experimental framework for the study, a preliminary test (pre-test) was conducted using two platforms: Lexica and Shutterstock. The objective was to assess the effectiveness of the prompts selected for the study. In this initial phase, six prompts were formulated combining one constant element with two variables. The constant element was the phrase “*smiling person*”, selected for the following connotations:

- Its singular form favours the representation of a single individual.
- The term does not imply a specific gender.
- The term does not imply a specific age.
- The term does not imply a specific ethnicity.
- The term does not imply any particular activity.
- It is expected that the person will be depicted frontally, despite no framing or compositional instructions being provided.

As variables, two differentiating elements were introduced. First, a spatial indicator was added to provide geographical context to the queries. Three alternative settings were selected: *city*, *landscape*, and *beach*.

The choice of the term *landscape* was based on a preliminary comparison between the use of *landscape* and *countryside*. As no significant differences were observed in the results returned by the platforms, *landscape* was selected as the more commonly used term.

Second, we introduced variations regarding the formal rendering of the images, with the aim of analysing how AI systems interpret the concept of photographic representation. Three for-

mulations were tested: *photography of*, *picture of*, and *image with photorealistic look*. This comparison seeks to examine not only how the systems interpret and represent degrees of realism or hyperrealism, but also the compositional and framing elements that emerge in the images generated or selected in response to each prompt.

The prompts used in the pre-test were as follows:

- Picture of a smiling person in a landscape.
- Photography of a smiling person in a landscape.
- Image of a smiling person in a landscape. Photorealistic look.

Similar tests were conducted for the two other settings, *city* and *beach*.

In all cases, prompts containing the term *photography* yielded more hyperrealistic results than those using *picture or image of (...) photorealistic look*. The latter option occasionally produced caricature-like images with clearly distorted features, characteristic of graphic illustration.

It was observed that incorporating the term *photography* in the prompt —added, as explained, to delimit the formal nature of the image in contrast to illustration— may introduce variations in the results, particularly when analysing the “activities” depicted in the scenes.

To assess the impact of this variable, additional tests were carried out using two comparable prompts: *Photography of a smiling person in the city* and *Smiling person in the city*, applied across the four platforms under analysis.

In the case of Shutterstock, including the term *photography* did indeed result in changes: several images featured people taking photographs, whereas such scenes were scarcely present when the term was omitted. By contrast, Getty Images returned virtually identical results in both cases, with no relevant variations attributable to the use of *photography*.

On the Lexica platform, as previously noted, the term *photography* functioned as an effective filter to exclude unwanted illustrations from the results. However, its inclusion or omission did not significantly alter any of the analytical parameters under consideration. Similarly, on Adobe Stock, no substantial changes in visual outcomes were observed in relation to the use of the term. As with Lexica, the absence of *photography* often led to the appearance of illustrations among the retrieved images.

In summary, it was found that the inclusion or exclusion of the term *photography* had no impact on the representation of biases linked to gender, race, age, or beauty-related stereotypes. Only in the case of Shutterstock was a slight deviation detected concerning the actions depicted, attributable to prompts explicitly including the term *photography*.

Consequently, the following three prompts were ultimately submitted to all four platforms:

- Photography of a smiling person in a landscape.
- Photography of a smiling person in the city.
- Photography of a smiling person on the beach.

These prompt combinations yielded the analytical sample, composed of a total of 600 images. The sample was generated by collecting the first 50 images returned by each of the four platforms, provided they met the established criteria and constituted a coherent response

to the three proposed queries. The size of this sample —600 images— has previously been employed in similar studies, such as that of Thurlow et al. (2020), which were likewise aimed at identifying biases and stereotypes in digital image banks.

## 4. What can be observed and how

This study is based on the hypothesis that prompts formulated without any explicit reference to the gender, age, or ethnicity of the individuals depicted should, in the absence of cultural or technical bias, yield a proportional distribution across the various observable categories. Any deviation from this theoretical balance is interpreted in the specialist literature as a potential indication of gender disparities and the presence of clichés or stereotypes (Cook & Cusack, 2010; Castillo-Mayén & Montes-Berges, 2014).

The overrepresentation of particular groups —whether in terms of gender, phenotype, or age range— may reveal visual repetitions: images that are similar in nature, featuring recurring compositional structures or iconic elements that reflect culturally accepted forms of coding. These, therefore, can be interpreted as stereotypes (Ángeles-Galiano, 2023).

In the field of media, such codified visual formulas, tropes, and stereotypes have proven to be highly effective communicative tools for illustrating events. They are part of professional routines in visual production and are especially prevalent in photographic imagery (Baeza, 2001; Freixa y Redondo-Arolas, 2022). Media outlets not only reproduce these visual constructs, but also continually update them through subtle variations, adapting them to the preferences and expectations of their audiences. In doing so, they contribute to the ongoing maintenance and symbolic renewal of these codes (Quin & McMahon, 1997; Novaes-Cirjanic, 2017).

The observation of formal and thematic repetitions, as well as the identification of common elements across the images analysed, facilitates the detection and description of the most recurrent stereotypes within the visual corpus. For this purpose, iconological methods are employed, and more specifically, iconographic description is used to identify those attributes potentially associated with visual biases stemming from socio-cultural clichés or stereotypes (Drainville, 2018; Hariman & Lucaites, 2007, 2016; Panofsky, 1979).

The application of a systematic analysis sheet is proposed as a methodological tool for assessing the presence of gender bias and stereotypes, structured around the parameters and indicators detailed in the following section.

Any image featuring more than one person, clearly depicting a group scene, or lacking a human figure altogether, was excluded from the analysis in order to ensure consistency and comparability of the results. Additionally, the textual information associated with each image was used as support to aid in its classification.

### 4.1. Parameters and indicators: Biases and stereotypes

Below is a description of the indicators proposed for observation, including the definition of each indicator, the observation procedure, and the values used.

#### 4.1.1. Indicator: Gender

- **Definition:** The prompts are designed to retrieve images that may depict one or more individuals. There is no explicit instruction specifying the gender of the person(s) represented.
- **Procedure:** If the image depicts a single individual, classify them as male, female, or undefined. Record the number of items representing each category.
- **Values:** Female / Male / Not defined.

#### 4.1.2. Indicator: Age

- **Definition:** The prompts are designed to retrieve images that may depict one or more individuals. There is no explicit instruction regarding the age of the person(s) represented.
- **Procedure:** If the image depicts a single individual, classify the age group represented. Record the number of items in each category.
- **Values:** Child / Young person / Adult / Older adult.

#### 4.1.3. Indicator: Ethnicity

- **Definition:** The prompts are designed to retrieve images that may depict one or more individuals. There is no explicit instruction regarding the ethnicity or ethnicities of the person(s) represented.
- **Procedure:** If the image depicts a single individual, classify the ethnicity represented. Record the number of items in each category.
- **Values:** Caucasian / Asian / African American / Latino-Mediterranean / Indigenous American / Middle Eastern.

#### 4.1.4. Indicator: Functional diversity

- **Definition:** The prompts are designed to retrieve images that may depict one or more individuals. There is no explicit instruction as to whether the individuals shown may or may not present any form of functional diversity.
- **Procedure:** If the image depicts a single individual, classify them according to different categories of functional diversity. Record the number of items in each category.
- **Values:** Motor disability / Sensory disability / Intellectual disability / No disability.

#### 4.1.5. Indicator: Beauty standards

- **Definition:** The prompts are designed to retrieve images that may depict one or more individuals. There is no explicit instruction as to whether the individuals shown conform to particular beauty standards or codes of visual representation.
- **Procedure:** If the image depicts a single individual, classify various visual elements that may be associated with beauty standards. The recurrence of such elements will allow for a shared labelling system. Record the number of items in each category.
- **A priori values:** None.
- **Possible values:** Make-up / Sunglasses / Glasses / Hats, caps, helmets / Jewellery / Watches / Summer clothing / Winter clothing / Backpacks / Handbags / Other.

#### 4.1.6. Indicator: Inventory of actions

- **Definition:** The prompts are designed to retrieve images that may depict one or more individuals in undefined settings. There is no explicit instruction as to whether the individuals should be performing any particular action. Actions may serve as indicators of pre-established and stereotyped behavioural codes.
- **Procedure:** If the image depicts a single individual, classify the actions being performed. The recurrence of elements will enable the establishment of a shared labelling system. Record the number of items in each category.
- **A priori values:** None, although a higher incidence of representations involving photographic devices is expected due to the nature of the prompt.
- **Possible values:** Taking a selfie / Looking at a map / Reading a book / Using a smartphone or tablet / Gesturing / Working / Eating or drinking / Using a camera or technology / Playing / Driving or being in a car / Dancing / Using public transport / Doing sport / Walking / Running / Other.

Once the results are obtained, comparisons are proposed for each parameter and indicator across the different platforms to confirm or rule out the presence of biases and stereotypes.

### 5. Expected results and limitations

The process of obtaining the sample is subject to particularities stemming from the algorithmic functioning of the platforms used. Despite the precision that prompts may convey, the systems do not always return images that strictly match the request. In this study, it was found that the result galleries frequently include group photographs, portraits of couples, or images in which no human figure appears at all.

Furthermore, traditional photographic stock libraries tend to offer multiple versions of the same scene. These variations typically present minimal differences, usually related to framing or composition. For instance, in response to a prompt requesting a portrait of a woman in an urban street, the system may return five or six virtually identical images, differing only slightly in framing or angle.

To construct the final sample, two types of images were removed: so-called false positives — that is, images not corresponding to the requested content— and redundant variations of the same shot, to avoid distortion in the analysis and to ensure greater visual diversity.

It was observed that image banks based on artificial intelligence tend to return more accurate results in response to user prompts. By way of illustration, the results showed that AI-based platforms achieved a 60.36% rate of prompt alignment, compared to 44.84% for conventional photographic stock banks. This difference can largely be attributed to the different business models underpinning each type of platform. While traditional banks are based on the exploitation of pre-existing archives —which leads them to offer available images even when they do not fully match the query— AI-driven platforms are designed to generate new content that more closely aligns with the user's request. As a result, they prioritise prompt accuracy over diversity of results.

The proposed methodological system has proven effective for observing the defined indicators. The results obtained (Freixa et al., 2025) reveal small but significant variations in the

perpetuation of bias when comparing the behaviour of AI-generated image platforms with that of traditional photographic banks.

This methodology seeks to provide a systematic tool to support the identification of bias and stereotypes, both in image banks and in any visual system that offers graphic responses to prompts formulated in natural language. In doing so, we propose an analytical framework that can be adopted by other research teams with the shared objective of making these dynamics visible and contributing to their correction. This task requires coordinated action across multiple dimensions—including those highlighted by our research—in order to move towards a more just and inclusive society.

## 6. Final note

Given its potential usefulness to other researchers and its applicability in other contexts, this chapter sets out the research methodology used in the study on gender bias and stereotypes in image banks, as presented in the following article:

Freixa, P., Redondo-Arolas, M., Codina, L., & Lopezosa, C. (2025). AI, Stock Photography, and Image Banks: Gender Biases and Stereotypes. *Hipertext.net*, (30), 197-214. <https://doi.org/10.31009/hipertext.net.2025.i30.05>

## 7. Funding

This work is part of the Project “Parameters and strategies to increase the relevance of media and digital communication in society: curation, visualisation and visibility (CUVICOM)”. Grant PID2021-123579OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU.

## 8. References

Abadal, E., & Codina, L. (2008). La diversidad cultural en Google y los motores de búsqueda: una aproximación conceptual. *Profesional de la Información*, 17(2), 191–198. <https://doi.org/10.3145/epi.2008.mar.09>

Aiello, G., Thurlow, C., & Portmann, L. (2023). Desocializing social media: The visual and media ideologies of stock photography. *Social media+society*, 9(1). <https://doi.org/10.1177/20563051231156363>

Allard, D. (August 8, 2023). Navigating the Intersection of AI and Journalism: the Future of Storytelling? *Kultur*. <https://www.kultur.art/navigating-the-intersection-of-ai-and-journalism-the-future-of-storytelling/>

Ángeles-Galiano, T. J. (2023). Análisis del estereotipo como imagen literaria: una propuesta de estudio. *Tropelías. Revista de Teoría de la Literatura y Literatura Comparada*, (39), 145-156. [https://doi.org/10.26754/ojs\\_tropelias/tropelias.2023397277](https://doi.org/10.26754/ojs_tropelias/tropelias.2023397277)

Baeza, P. (2001). *Por una función crítica de la fotografía de prensa*. Gustavo Gili.

Battelle, J. (2006). *Buscar: cómo google y sus rivales han revolucionado los mercados y transformado nuestra cultura*. Urano.



- Bright, J. (20 de noviembre 2023). Photography & AI. *The independent photographer*. <https://independent-photo.com/news/photography-ai/>
- Castillo-Mayén, R., & Montes-Berges, B. (2014). Análisis de los estereotipos de género actuales. *Anales de Psicología*, 30(3), 1044-1060. <https://www.redalyc.org/pdf/167/16731690027.pdf>
- Codina, L. (2011). Entender los bancos de imágenes. *Profesional de la Información*, 20(4), 417-423. <https://doi.org/10.3145/epi.2011.jul.08>
- Codina, L. (2018). Sistemas de búsqueda y obtención de información: componentes y evolución. *Anuario ThinkEPI*, 12, 77-82. <https://doi.org/10.3145/thinkepi.2018.06>
- Codina, L., & Lopezosa, C. (2020). *Documentación periodística: búsqueda de imágenes en Internet y bancos de fotografía y vídeo 2020*. Universitat Pompeu Fabra. <https://hdl.handle.net/10230/44063>
- Cook, R. J., & Cusack, S. (2010). *Estereotipos de género. Perspectivas legales transnacionales*. Profamilia. [https://www.law.utoronto.ca/utfl\\_file/count/documents/reprohealth/estereotipos-de-genero.pdf](https://www.law.utoronto.ca/utfl_file/count/documents/reprohealth/estereotipos-de-genero.pdf)
- Drainville, R. (2018). Iconography for the Age of Social Media. *Humanities*, 7(1), 12. <https://doi.org/10.3390/h7010012>
- Ferry, S. (2 de mayo 2023). Fotógrafo humano versus Inteligencia Artificial. *Fundación Gabo*. <https://fundaciongabo.org/es/etica-periodistica/blogs/fotografo-humano-versus-inteligencia-artificial>
- Freixa, P., & Redondo-Arolas, M. (2022). Mirando al mar. Tropos visuales de resiliencia y superación en la COVID-19. *Arte, Individuo y Sociedad*, 34(2), 777-798. <https://doi.org/10.5209/aris.75517>
- Freixa, P., & Redondo-Arolas, M. (2023). IA y fotoperiodismo: un escenario por definir. In *XV Congreso Internacional de Comunicación Digital y Ciberperiodismo. Desinformación en la era de la inteligencia artificial*. 27 y 28 de noviembre de 2023 UPV/EHU. <https://hdl.handle.net/10230/68388>
- Freixa, P., & Redondo-Arolas, M. (2024). IA, contenido visual y periodismo: fundamentos para un debate. *IX Congreso Internacional de la AE-IC: Comunicación e Innovación Sostenible*. Murcia, 28 mayo - 1 junio 2024.
- Freixa, P., Redondo-Arolas, M., Codina, L., & Lopezosa, C. (2025). AI, Stock Photography, and image banks: Gender biases and stereotypes. *Hipertext.net*, (30), 197-214. <https://doi.org/10.31009/hipertext.net.2025.i30.05>
- Frosh, P. (2003). *The image factory: Consumer culture, photography and the visual content industry*. Berg Publishers.
- Frosh, P. (2015). Beyond the image bank: Digital Commercial Photography . En: M. Lister (Ed.), *The photographic image in digital culture* (pp. 131-148). Routledge.
- Frosh, P. (2020). Is commercial photography a public evil? Beyond the critique of stock photography. In *Photography and its publics* (pp. 187-206). Routledge.



- Goodnow, T. (2010). Visual bias in time's "the great divide": A semiotic analysis of Clinton and Obama photographs. *American Behavioral Scientist*, 54(4), 406–416. <https://doi.org/10.1177/0002764210381865>
- Gynnild, A. (2017). The Visual Power of News Agencies. *Nordicom Review*, 38(2), 25-39. <https://doi.org/10.1515/nor-2017-0412>
- Hariman, R., & Lucaites, J. L. (2007). *No caption needed: Iconic photographs, public culture, and liberal democracy*. University of Chicago Press
- Hariman, R., & Lucaites, J. L. (2016). *The public image: Photography and civic spectatorship*. University of Chicago Press.
- Hughes, C. (January 1, 2024). AI-generated images vs. Stock Photography: A comprehensive comparison. *BlogHerald*. <https://blogherald.com/seo/ai-generated-images-vs-stock-photography-a-comprehensive-comparison/> [discontinued entry]
- Kamin, D. (2023). *Picture-work: How libraries, museums, and stock agencies launched a new image economy*. MIT Press.
- Kwak, H., & An, J. (2021). Revealing the hidden patterns of news photos: Analysis of millions of news photos through GDELT and deep learning-based Vision APIs. *Proceedings of the International AAAI Conference on Web and Social Media*, 10(2), 99-107. <https://doi.org/10.1609/icwsm.v10i2.14840>
- Luebke, B. F. (1989). Out of focus: Images of women and men in newspaper photographs. *Sex Roles*, 20(3-4), 121–133. <https://doi.org/10.1007/BF00287986>
- Machin, D. (2004). Building the world's visual language: The increasing global importance of image banks in corporate media. *Visual Communication*, 3(3), 316-336. <https://doi.org/10.1177/1470357204045785>
- Machin D., & Polzer L. (2015). *Visual Journalism*. Palgrave Macmillan.
- May, T. (2024, March 19). The best stock photo sites in 2024. *Digital Camera World*. <https://www.digitalcameraworld.com/buying-guides/best-stock-photo-sites> [discontinued entry]
- Miller, S. H. (1975). The content of news photos: Women's and men's roles. *Journalism and Mass Communication Quarterly*, 52(1), 70. <https://doi.org/10.1177/107769907505200112>
- Mortensen, T. M., & Gade, P. J. (2023). Does photojournalism matter? News image content and presentation in the middletown (NY) Times Herald-record before and after layoffs of the photojournalism staff. *Journalism & Mass Communication Quarterly*, 95(4), 990-1010. <https://doi.org/10.1177/1077699018760771>
- Mortensen, T. M., McDermott, B. P., & Ejaz, K. (2023). Measuring photo credibility in journalistic contexts: Scale development and application to staff and stock photography. *Journalism Practice*, 17(6), 1158-1177. <https://doi.org/10.1080/17512786.2021.1976073>
- Mortensen, T. M., Piacentine, C., Wen, T., Bost, N., & McDermott, B. P. (2024). True or false? Viewer perceptions of emotional staff and stock photos in the news. *Journal of Media Ethics*, 39(1), 16–32. <https://doi.org/10.1080/23736992.2023.2295942>

- Novaes-Cirjanic, J. (2017). *Aproximación deontológica al fotoperiodismo: Creación de estereotipos en la visión occidental del "tercer mundo"* [Tesis Doctoral, Universitat Autònoma de Barcelona]. <https://ddd.uab.cat/record/187161>
- Panofsky, E. (1979). *El significado en las artes visuales*. Alianza Editorial
- Perdices-Castillo, L., & Perianes-Rodríguez, A. (2011). Sistemas de búsqueda y visualización en bancos de imágenes comerciales. *Profesional de la información*, 20(4), 439–443. <https://doi.org/10.3145/epi.2011.jul.11>
- Piironen, J. (5 de noviembre 2022). Side-by-Side Comparison of The Most Popular Sites To Sell Stock Photos. *Jamo Images*. <https://jamoimages.com/side-by-side-comparison-of-the-most-popular-sites-to-sell-stock-photos/>
- Piironen, J. (12 de noviembre 2023). The 3 best websites to sell stock photos and a bonus. *Jamo Images*. <https://jamoimages.com/the-3-best-websites-to-sell-stock-photos/>
- Quin, R., & McMahon, B. (1997). *Historias y estereotipos*. Grafo.
- Reporters Sans Frontières, RSF. (10 de Noviembre 2023). RSF and 16 partners unveil Paris Charter on AI and Journalism. <https://rsf.org/en/rsf-and-16-partners-unveil-paris-charter-ai-and-journalism>
- Rodgers, S., & Thorson, E. (2000). "Fixing"; stereotypes in news photos: A synergistic approach with the *Los Angeles Times*. *Visual Communication Quarterly*, 7(3), 8–7. <https://doi.org/10.1080/15551390009363436>
- Rodgers, S., Kenix, L. J., & Thorson, E. (2007). Stereotypical portrayals of emotionality in news photos. *Mass Communication & Society*, 10(1), 119–138. <https://doi.org/10.1080/15205430709337007>
- Rönnback, R., Emmery, C., & Brighton, H. (2025). Automatic large-scale political bias detection of news outlets. *PLoS One*, 20(5), e0321418. <https://doi.org/10.1371/journal.pone.0321418>
- Sojit-Pejcha, C., & Crabapple, M. (11 mayo 2023). Artists and journalists call for the restriction of AI illustration in publishing. *Document Journal*. <https://www.documentjournal.com/2023/05/ai-art-generators-molly-crabapple-open-letter-copyright-law-stability-ai-midjourney-dall-e>
- Thurlow, C., Aiello, G., & Portmann, L. (2020). Visualizing teens and technology: A social semiotic analysis of stock photography and news media imagery. *New Media & Society*, 22(3), 528–549. <https://doi.org/10.1177/1461444819867318>
- Tsang, K. J. (1984). News photos in *Time* and *Newsweek*. *Journalism Quarterly*, 61(3), 578–723. <https://doi.org/10.1177/107769908406100314>
- Vrabič-Dežman, D. (2024). Promising the future, encoding the past: AI hype and public media imagery. *AI and Ethics*, 4, 743–756. <https://doi.org/10.1007/s43681-024-00474-x>
- Waldman, P., & Devitt, J. (1998). Newspaper photographs and the 1996 presidential election: The question of bias. *Journalism & Mass Communication Quarterly*, 75(2), 302–311. <https://doi.org/10.1177/107769909807500206>

# Critical thinking and artificial intelligence in academia: A qualitative matrix analysis procedure for evaluating AI systems

**Lluís Codina**

Universitat Pompeu Fabra, Spain

<https://orcid.org/0000-0001-7020-1631>

**Elisenda Aguilera-Cora**

Universitat Pompeu Fabra, Spain

<https://orcid.org/0000-0003-0923-9192>

**Carlos Lopezosa**

Universitat de Barcelona, Spain

<https://orcid.org/0000-0001-8619-2194>

**Pere Freixa**

Universitat Pompeu Fabra, Spain

<https://orcid.org/0000-0002-9199-1270>

---

Codina, L., Aguilera-Cora, E., Lopezosa, C., & Freixa, P. (2025). Critical thinking and artificial intelligence in academia: A qualitative matrix analysis procedure for evaluating AI systems. In J. Guallar, M. Vázquez, & A. Ventura-Cisquella (Coords). *Digital communication. Trends and good practices* (pp. 161-173). Ediciones Profesionales de la Información. <https://doi.org/10.3145/cuvicom.12.eng>

---

## Abstract

This work introduces the Matrix AI Systems Analysis Procedure (MASIA), a qualitative, matrix-based method designed to evaluate the performance and quality of generative artificial intelligence systems within academic settings. MASIA centers on the analysis of three key components in AI-generated responses: narrative synthesis, source usage, and the formulation of new prompts. By doing so, it fosters critical thinking among users and offers valuable tools for both teaching and research. The procedure defines variables and analytical parameters that enable the comparison of different AI systems, thereby supporting informed decision-making in scholarly and research environments. Furthermore, MASIA integrates ethical considerations, including traceability, proper attribution, and plagiarism prevention, making it a flexible instrument adaptable to various academic needs and projects. The chapter concludes that MASIA is a straightforward yet powerful tool for enhancing critical thinking, improving teaching and learning processes, and providing a foundation for comparative research on artificial intelligence in academia.

## Keywords

Generative artificial intelligence; Qualitative evaluation; Critical thinking; Analysis matrices; Academic ethics; AI systems in academy; Evaluative methods.

---

## 1. Introduction

This paper presents an analytical procedure to evaluate the performance and quality of generative artificial intelligence systems in academic environments.

The procedure, which we call *the Matrix AI Systems Analysis Procedure* or MASIA, is designed to evaluate AI systems that, as part of their response, not only provide a narrative summary but also include citations and the bibliographic sources they used to generate the content.

This method of analysis promotes critical thinking among AI system users, provides elements for teaching-learning processes, and can be the basis for developing data collection in research processes.

The use of sources as part of the response is a necessity in the academic context because one can verify and expand the information provided by the AI, as well as maintain the chain of attributions (High-Level Expert Group on Artificial Intelligence . 2019; Crompton and Burke, 2023; Kaebnick et al. 2023; Lund et al, 2023; Tilie et al., 2023; Gundersen et al., 2018; World Commission on the Ethics of Scientific Knowledge and Technology, 2019; Dwivedi et al., 2021; Bianchini et al. 2022). The latter is doubly convenient, because in addition to increasing the quality of the AI response, it prevents plagiarism or inadequate attribution, both of which are essential in academic work.

The procedure presented here consists of analysis matrices based on a series of variables (Codina & Pedraza, 2016), which are determined according to the authors' teaching experience and prior experience in analyzing and using AI systems for academic work, particularly due to the need to provide protocols for the critical use of AI systems to university students and predoctoral researchers.

Since the launch of ChatGPT in late 2022, the authors have incorporated artificial intelligence into their teaching and research activities (Lopezosa & Codina, 2023; Lopezosa et al., 2023a; 2023b; Aguilera-Cora et al., 2024a; 2024b; Codina, 2025). This integration highlighted the need for an intellectual tool to train students and predoctoral researchers in the proper use of AI (Codina and Garde, 2023). There was also a need for an instrument that would allow comparative studies of the efficiency of artificial intelligence systems suitable for use in academia. The latter could be useful for research purposes, or for providing economic decision-makers at universities, for example, with information on which to evaluate the acquisition of AI systems (Bhatia, 2023; Whitfield & Hofmann, 2024; Elsevier 2024).

Before presenting the components of the evaluation method, we must present some terminological clarifications, the consideration of which is part of the procedure itself, just as we must consider the composition of the results of AI when it responds to a user instruction.

## 2. Terminology

The terminology presented in the table below is considered part of the procedure, so it is necessary to precisely establish the use of a set of terms for its proper application.

**Table 1**  
*Terminology of the AI systems evaluation procedure*

Term	Explanation
Bibliographic sources	In a RAG-type AI system (see definition below), this is the list of documents (journal articles, reports, web pages, etc.) that justify the answer. In the context of the evaluation procedure, this concept of source is the one taken into account unless otherwise indicated.
Sources of information	In a RAG-type AI system (see definition below), information sources are the resources used to locate the sources on which its response is based. Typical information sources for RAG-type AI systems can be academic databases or search engines like Google.
Indicators	In an evaluation procedure, indicators are characteristics that provide information about what is to be evaluated or compared. For example, in a comparative analysis of national economies, the unemployment rate, inflation, or GDP are indicators. By their very nature, they are also variables.
Matrices	Table-based information structures that allow data to be extracted, presented visually, and comparatively analysed. In the procedure presented here, the use of matrices is considered normative. The word "table" is equivalent. The use of the term "matrix" introduces the idea that a table meets certain conditions, the most important of which are that they are homogeneous tables and are organized so that rows are entities and columns are the properties of the entities.
AI model	Also called Large Language Model (LLM). This is the technical name for generative artificial intelligence, given its technological foundation. An AI model or LLM cannot be used by an end user, as its use requires programming and APIs. Therefore, end users generally work with AI through AI systems.
Results page	An AI system's response is presented to the user on a results page that typically consists of at least three components: the narrative summary, a list of sources, and a set of suggested new prompts.
Parameters	In an evaluation procedure, parameters group indicators or variables. The idea is that a group of variables serves to characterize a significant aspect of a certain complexity of the subject being evaluated. For example, one parameter of nations in comparative analyses is their economy, another their demographics or political system, etc. But to characterize each parameter, it is necessary to use disaggregated indicators, such as the unemployment rate in the case of the economy, along with others such as GDP, etc.
Prompt	Natural language instruction used to obtain the response from an AI system.

Term	Explanation
Suggested Prompts	List of new prompts that some AI systems provide as part of their response to a prompt.
Rapid review	A type of review that omits some of the usual controls of systematic reviews to obtain immediate results with preliminary value. Narrative syntheses from AI can be compared to a form of rapid review.
Retrieval Augmented Generation (RAG)	Augmented Retrieval (ARG) involves improving the answers generated by AI systems by combining their training base knowledge with information retrieved in real time from external sources such as academic or specialized databases or general-purpose search engines like Google or Bing. Most academic AI systems are ARG. Some general-purpose AI systems, like Perplexity, are also ARG-type systems. Google, presumably, should become ARG-type once it effectively integrates its search engine with its AI.
Literature review	A literature review is a systematic process of searching, selecting, analyzing, and synthesizing existing information on a specific topic. It involves critically evaluating previous studies, identifying patterns, debates, and gaps in current knowledge, and presenting a comprehensive and organized overview of the state of the art in the field. , Producing literature reviews applied to academia is one of AI's main functions.
Scratchpad	The scratchpad is the section preceding the narrative synthesis where the AI presents the chain of reasoning it followed to accomplish its tasks. In our case, the task consists of solving the four phases leading to a narrative synthesis. The ability to examine the scratchpad means, among other things, checking whether the AI understood the objectives of the task and correctly approached it.
Narrative synthesis	A narrative synthesis is the result of analyzing a set of sources using a well-defined framework and compiling the key insights derived from the analysis of these documents into a coherent textual summary. A narrative synthesis is both a byproduct of a literature review and part of the response of a generative AI system.
AI system	It is composed of one or more AI models (also called LLMs for Large Language Model), of one or more software layers, e.g. , for querying databases (e.g. , academic databases), for managing references, etc., as well as a user interface.
Utilities	These consist of complementary functions that software programs typically offer in addition to their core functions. For example, in an AI system focused on academia, a utility might consist of resources for managing references.
Variables	A variable is a property of an entity that can take on different values for each entity, or over time within the same entity. Recording these values allows entities to be analyzed, characterized, and compared. They can also be called indicators if their heuristic nature is desired.

Source: own elaboration

### 3. Composition of an AI system's results page

In a results page of the type of AI system we are analyzing, we can determine the existence of three main components:

1. Narrative synthesis.
2. Sources consulted.
3. New prompts.

We will present each of these components for evaluation purposes. But first, we must point out a fourth element that, although not part of the results page, can be found within the interface itself:

4. Additional utilities and functionalities specific to each system.



### 3.1. Narrative synthesis

Narrative synthesis is the text generated by generative artificial intelligence. It's a synthesis because it's the result of analyzing and synthesizing a series of previous pieces of information. It's narrative because it's presented as a more or less articulated narrative or discourse. Other types of syntheses are possible, such as those presented in the form of tables or graphs. However, when we talk about narrative syntheses, we consider any format to be included, by extension, unless otherwise stated.

From here, we can establish a first block of criteria according to which AIs that present synthesis are preferable, which are:

- **Traceable through a visible *scratchpad*.** The *scratchpad* is a section preceding the narrative summary in which the AI transparently and traceably presents the chain of reasoning it followed to solve the task. The ability to examine this chain of reasoning allows for the detection of potential bias or other errors, as well as verifying whether the AI correctly understood the task. In any case, it facilitates the traceability of the process followed.
- **Articulated.** This means that the summary is presented in some sort of structure. For example, in separate sections, possibly organized by headings and following some sort of logical arrangement of sections.
- **They exhibit coherence and cohesion.** Coherence consists of the interrelationship of the sentences that make up the text through their appropriate connection to the main theme. It is manifested by thematic unity, the (relative) absence of redundancy, and the logical progression of ideas. Cohesion, on the other hand, is manifested in the grammatical interrelationship of sentences. It is primarily determined by the connectives.
- **They exhibit connectivity.** The end of each paragraph anticipates the next, and the beginnings of subsequent paragraphs connect to the previous ones. This property is made evident through the use of connectives. Connectivity is increased if there is a section that reunites the main ideas, or a section with an equivalent function.
- **They are (relatively) long.** All other criteria being equal, long summaries are preferable. Since we're talking about a range that can extend from 200 to 3,000 words, those that, if necessary, can be closer to this higher limit are preferable.
- **They are multimodal.** In addition to text, they include some additional formatting, e.g., tables, cards, concept maps, mind maps, or diagrams.

### 3.2. Sources

In academic AI, sources are typically documents, reports, and scientific journal articles. Ideally they allow the ideas and content comprising the synthesis to be attributed to their original creators. In the academic context, we have a categorical imperative to use AI systems that, along with the generated narrative synthesis, provide the sources on which they are based. This is the reason for preferring RAG-type AI systems.

An additional reason for this preference is that an absence of sources in the answer would lead to a break in the attribution chain and, consequently, predispose to plagiarism. Note that we are separating the actual act (or lack thereof) of plagiarism from the fact that some AI systems facilitate or promote, de facto, plagiarism by offering unsourced answers. Logically, we should prefer AI systems that, at the very least, do not promote plagiarism.



For the purposes of our proposed analysis, AIs are preferable [to what?] in relation to sources:

- They exhibit **capillarity**. AIs that assign sources at the sentence level, or failing that, at the paragraph level, are preferable to a final list that affects the entire undifferentiated narrative synthesis. Capillarity also implies connectivity, since when a paragraph has (for example) three related ideas, each source is linked to each of the ideas, instead of placing the three sources at the end of the paragraph or at the end of the entire synthesis.
- They provide well-formed citation **formats**, that is, with complete reference information and, where appropriate, viable links.

Furthermore, the user of AI is obliged to verify and review sources, not only to evaluate arguments, but also to attribute third-party ideas and content to their true authors through the conventional citation system.

### 3.3. New prompts

Some AIs offer, as a third prominent component of their responses, a list of new prompts or new questions. This approach may be of little interest or may be very incisive. In the latter case, they have obvious heuristic value. Preferable are those that generate additional, related prompts or questions as part of the results page, which we will evaluate:

- **Opportunity**. That is, are the new suggested prompts appropriate for the search objectives?
- **Variety of approaches**. Do they offer new facets or approaches not considered in the original prompt?

### 3.4. Utilities and *Idiofunctions*

Some AI systems present one or more characteristics that are specific and unique to the system under consideration. In contrast to common functions, we can speak of *idiofunctions*; that is, functions unique to each particular system and therefore present only in the system under consideration. For example, an AI system may present a function that consists of extracting concepts, or another that consists of being able to design analysis matrices from references.

These are called *idiofunctions* because they are unique to each AI. Although these functions may become standardized over time (and the concept may lose its meaning), these differences are significant at present and useful to consider.

### 3.5. Analysis matrices

With the help of the previous concepts, we can now present the elements of analysis, which we articulate in parameters and variables as shown in Table 2.

Table 2  
Analysis variables.

Parameter	Code	Variables / Check Question
<b>1. Narrative synthesis</b>	1.1	<b>Scratchpad</b> Does it present a scratchpad with the chain of reasoning followed by the AI system? Can this document be consulted after the task is completed?
	1.2	<b>Articulation</b> Is the narrative synthesis presented organized or articulated in various sections or is it presented as a continuum without a defined structure?
	1.3	<b>Coherence and connection</b> Is there a consistent thematic unity throughout the narrative summary and within each paragraph? Is there a connection between the sections, paragraphs, or sections of the narrative summary?
	1.4	<b>Extension</b> Is the summary of the narrative adequate for its objectives? How many words does the narrative summary contain? Are there alternative versions of the extension?
	1.5	<b>Multimodality</b> Does the results page include only text, or does it include other forms of information, such as diagrams? If not initially presented, are they offered as alternatives?
<b>2. Sources</b>	2.1	<b>Number</b> How many sources are cited?
	2.2	<b>Diversity</b> Do the sources exhibit adequate diversity for the purposes of the prompt? Note: A single database is not an <i>a priori</i> limitation on diversity. Does the system allow you to differentiate whether the sources belong to academic texts, press, grey literature, or other unregulated sources?
	2.3	<b>Capillarity</b> Are the sources connected at least at the paragraph or section level?
	2.4	<b>Well formed</b> Are the sources presented in a format that is easy to export, manage, and cite sources?
<b>3. Suggested prompts</b>	3.1	<b>Chance</b> Do the suggested new prompts seem appropriate or timely given the information needs?
	3.2	<b>Variety</b> Are the prompts varied and help broaden the focus of the topic?
<b>4. Idiofunctions</b>	4.1	<b>Specific and exclusive functions of each system considered</b> In addition to the common functions examined, does the system have any other specific functions?

**Table 3**  
*Theoretical scores.*

Parameter	Code	Variables	Theoretical score
<b>Narrative synthesis</b>	1.1	Scratchpad	0-3
	1.2	Joint	
	1.3	Connection	
	1.4	Extension	
	1.5	Multimodality	
<b>Sources</b>	2.1	Number	
	2.2	Diversity	
	2.3	Capillarity	
	2.4	Well formed	
<b>Suggested Prompts</b>	3.1	Chance	
	3.2	Variety	
<b>Idiofunctions</b>	4.1	Specific and exclusive functions of each system considered	0-3

The scoring scale is offered as an example. For each use, those responsible may (with justification) determine other scales. In this case, we have used a scale typical of heuristic evaluations in the field of information systems usability, and it corresponds to the following estimate:

Punctuation	Interpretation
0	Absence of function or variable considered
1	The function or variable appears in a minimal expression
2	The function or variable is correctly implemented but allows for improvements
3	The function or variable is fully implemented

The initial scores in this scoring system are assigned intuitively and are adjusted as new cases are examined to allow comparisons. A final tally is made once all cases are examined. It is also common for two analysts to assign scores independently, then the scores are compared, and discrepancies are resolved by consensus. However, the scale and the specific procedure for assigning scores can be established for each specific project.

**Table 4**  
*Data extraction table*

Cod.	Variable	Punctuation
	<b>Narrative synthesis</b>	
1.1	Scratchpad	
1.2	Joint	
1.3	Connection	
1.4	Extension	
1.5	Multimodality	
	<b>Sources</b>	
2.1	Number	
2.2	Diversity	
2.3	Capillarity	
2.4	Well formed	

Cod.	Variable	Punctuation
	<b>Additional Prompts</b>	
3.1	Chance	
3.2	Variety	
	<b>Idiofunctions</b>	
4.1	Specific and exclusive functions of each system considered	
<b>TOTAL</b>		

Comparative summary table

System	Synthesis	Sources	Prompts	Idiofuncion	TOTAL

The tables above are common examples of matrix-based analysis systems. For each specific project, project managers can modify any aspects as appropriate.

### 3.6. Other evaluation modes

It is clear that different evaluation methods can be developed. Task-based evaluation is one significant alternative, as in Font-Julián et al. (2024), in which an evaluation model is developed that combines qualitative and quantitative analysis procedures under specific tasks and with a group of two or more users as judges who agree on their evaluations.

Another significant alternative are the *benchmarking* evaluation methods, such as those that can be seen on the *Artificial Analysis portal* (<https://artificialanalysis.ai/>) where dozens of language models are periodically compared based on a battery of tests.

### 3.7. Differential contribution of this evaluation mode

Any evaluation method can be useful, depending on the context and objectives of each case. The qualitative evaluation method we propose here has a threefold function:

- Strengthen users' critical thinking regarding AI.
- Provide a method for teaching/learning and acquiring skills in the use of AI systems.
- Provide a procedure for evaluating and comparatively analyzing AI systems based on qualitative matrix analysis of parameters and indicators.

### 3.8. Variable geometry procedure

The MASIA procedure provides analytical frameworks that can be applied as presented. However, it is not essential to use all variables, and new variables can be added or even other parameters can be considered. The essence of this evaluation method is as follows:

- The designers of the analysis, with any of the three objectives stated above, may consider the convenience of adding new variables or removing some of the variables.

- Based on the response structure and utilities of an AI system, a series of analysis variables are proposed, for the evaluation of which a concise but minimally viable guide is available.
- The analyses are carried out using matrices such as those shown in the preceding tables, which facilitate both data collection and comparison.

If the analysis is being conducted for research purposes, the team wishing to apply the MASIA procedure must add the necessary validity, reliability, transparency, and traceability protocols that are common in high-quality research.

### 3.9. Phases of the procedure

We conceive the MASIA procedure as part of a workflow with three distinct phases. One phase is specific to the procedure presented in this chapter, and two additional phases that should always be taken into account when using AI systems in academic contexts, regardless of the specifics that determine their use. [would be helpful to say what those other two phases are here – there are three headings below so it's not easy to follow which phases are meant].

#### 3.10. Preparation

This is the initial phase, where the researcher determines whether it is appropriate to use AI systems to solve a task, as well as select the most appropriate one. In this first phase, the researcher must consider the objective of the task to be solved, as well as the different AI systems that can best contribute to its achievement, if applicable, whether through free or subscription-based tools, proprietary or institutional. Aspects related to privacy and data protection must also be considered. Finally, it is advisable to conduct a test or pretest before formally using an AI system, especially when using a given AI system for the first time, to facilitate an initial assessment of the tool's effectiveness in an area the researcher is familiar with. This phase is recommended for any model that integrates the use of AI systems in academic contexts.

#### 3.11. Analysis

This phase involves the MASIA procedure for evaluating AI systems that present, on their results page, at least the following three elements: narrative summary, sources consulted, and new prompts. The researcher evaluates different AI systems using matrices and variables, assigning scores based on the completion of a series of checklist questions, which allow them to identify which AI system is the most suitable.

#### 3.12. Research ethics

This phase should guide the entire process, from the conception of the task or research project for which AI systems will be used, to its completion. As in the preparation phase, research ethics should be part of any research project that uses AI systems, regardless of the approach or model adopted. The research ethics phase includes documenting the process, recording the prompts used and the date and context of interaction, or taking screenshots to promote transparency and reproducibility. Results should also be validated, and the sources

obtained in the different systems used must be verified. Another important aspect is avoiding plagiarism, both in responses based on previous work and in content generated by the AI systems themselves. Furthermore, in academic contexts, it is essential to maintain the traceability of original work and properly cite the use of AI systems.

## 4. Conclusions

We have presented a relatively simple procedure for analyzing and comparing AI systems applied to academic work and settings. The procedure covers several functions, the first and perhaps most notable of which is that it supports critical thinking among AI users by enabling them to evaluate significant aspects of AI.

Two other functions are that, for the reason stated above (but not reducible to it), the procedure can be used in teaching-learning processes in the training of university students and predoctoral researchers.

Finally, this procedure can be used for research purposes. Team members have systematic frameworks to analyze and compare various AIs based on a single prompt or a small set of prompts.

In general, we understand that this procedure, as we noted above, is easy to interpret and apply, without prejudice to its heuristic capacity, which can be significant thanks to the effort of analysts.

## 5. Funding

This work is part of the Project "Parameters and strategies to increase the relevance of media and digital communication in society: curation, visualisation and visibility (CUVICOM)". Grant PID2021-123579OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU.

## 6. References

Aguilera-Cora, E., Lopezosa, C., & Codina, L. (2024a). *Scopus AI beta: Functional analysis and cases* (DigiDoc Editorial Series. DigiDoc Reports). Pompeu Fabra University, Department of Communication. <https://repositori.upf.edu/handle/10230/58658>

Aguilera-Cora, E., Lopezosa, C., Fernández-Cavia, J., & Codina, L. (2024b). Accelerating research processes with Scopus AI: A place branding case study. *Pan-American Journal of Communication*, 6(1). <https://doi.org/10.21555/rpc.v6i1.3088>

Bianchini, S., Müller, M., & Pelletier, P. (2022). Artificial intelligence in science: An emerging general method of invention. *Research Policy*, 51(10), 104604. <https://doi.org/10.1016/j.respol.2022.104604>

Bhatia, P. (2023). ChatGPT for academic writing: A game changer or a disruptive tool? *Journal of Anaesthesiology Clinical Pharmacology*, 39(1), 1. [https://doi.org/10.4103/joacp.joacp\\_84\\_23](https://doi.org/10.4103/joacp.joacp_84_23)

Codina, L. (2025). *Sistemas de inteligencia artificial para la síntesis del conocimiento: marcos teóricos y revisiones de la literatura*. [Presentación]. Facultad de Comunicación, Universidad Pompeu Fabra. <http://hdl.handle.net/10230/70242>

Codina, L., & Garde, C. (2023). *Uso de ChatGPT en la docencia universitaria: fundamentos y propuestas* [Informe]. Facultad de Comunicación, Universidad Pompeu Fabra. <http://hdl.handle.net/10230/570157>

Codina, L., & Pedraza-Jiménez, R. (2016). Características y componentes de un sistema de análisis de medios digitales: el SAAMD. In R. Pedraza-Jiménez, L. Codina, & J. Guallar (Eds.), *Calidad en sitios web: método de análisis general, e-commerce, imágenes, hemeroteca y turismo* (pp. 15–40). UOC Publicaciones.

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20(1). <https://doi.org/10.1186/s41239-023-00392-8>

Dwivedi, Y.K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J.S., Eirug, A., Galanos, V., Ilavarasan, P.V., Janssen, M., Jones, P., Kar, A.K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., ... Williams, M.D. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 57, 101994. <https://doi.org/10.1016/j.ijinfomgt.2019.08.002>

Elsevier. (2024, January 16). *Launch of Scopus AI to help researchers navigate the world of research*. <https://www.elsevier.com/about/press-releases/launch-of-scopus-ai-to-help-researchers-navigate-the-world-of-research>

Font-Julián, C. I., Orduña-Malea, E., & Codina, L. (2024). ChatGPT Search as a tool for scholarly tasks: Evolution or return? *Infonomy*, 2(5). <https://doi.org/10.3145/infonomy.24.059>

Gundersen, O. E., Gil, Y., & Aha, D.W. (2018). On reproducible AI: Towards reproducible research, open science, and digital scholarship in AI publications. *AI Magazine*, 39(3), 56–68. <https://doi.org/10.1609/aimag.v39i3.2816>

High-Level Expert Group on Artificial Intelligence. (2019, April 8). *Ethics guidelines for trustworthy AI*. European Commission. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Kaebnick, G. E., Magnus, D., Kao, A., Hosseini, M., Resnik, D. B., Dubljević, V., Rentmeester, C. A., Gordijn, B., & Cherry, M. J. (2023). Editor's statement on the responsible use of generative AI technologies in scholarly journal publishing. *Medicine, Health Care and Philosophy*, 26(4), 499–503. <https://doi.org/10.1007/s11019-023-10176-6>

Lopezosa, C., & Codina, L. (2023). Testing Bard: Probandando Bard: así funciona la Inteligencia Artificial Generativa de Google. *Anuario ThinkEPI*, 17. <https://doi.org/10.3145/thinkepi.2023.e17a25>

Lopezosa, C., Codina, L., & Ferran-Ferrer, N. (2023a). *ChatGPT como apoyo a las systematic scoping reviews: integrando la inteligencia artificial con el framework SALSA*. Col·lecció del CRICC. Barcelona: Universitat de Barcelona.

Lopezosa, C., Codina, L., & Boté-Vericad, J. J. (2023b). *Testeando ATLAS.ti con OpenAI: hacia un nuevo paradigma para el análisis cualitativo de entrevistas con inteligencia artificial*. Barcelona: Universitat Pompeu Fabra. Departament de Comunicació. [30] p. (Serie Editorial DigiDoc. DigiDoc Reports).



Lund, B., Wang, T., Mannuru, N. R., Nie, B., Shimray, S. R., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial intelligence-written research papers and the ethics of large language models in scholarly publishing. *Journal of the Association for Information Science and Technology*, 74(5), 570–581. <https://doi.org/10.1002/asi.24750>

Tlili, A., Huang, R., Mustafa, M. Y., Zhao, J., Bozkurt, A., Xu, L., ... & Wang, H. (2023). Speaking of transparency: Are all artificial intelligence (AI) literature reviews in education transparent? *The Journal of Applied Learning & Teaching (JALT)*, 6(2). <https://doi.org/10.37074/jalt.2023.6.2.15>

UNESCO. (2023, September 8). *Guidance for generative AI in education and research*. <https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research>

Whitfield, S., & Hofmann, M. A. (2023). Elicit: AI literature review research assistant. *Public Services Quarterly*, 19(3), 201–207. <https://doi.org/10.1080/15228959.2023.2224125>

World Commission on the Ethics of Scientific Knowledge and Technology. (2019) *Preliminary study on the ethics of artificial intelligence*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000367823>

# Web of Science Research Assistant: Functional analysis and usage recommendations

**Carlos Lopezosa**

Universitat de Barcelona, Spain

<https://orcid.org/0000-0001-8619-2194>

**Elisenda Aguilera-Cora**

Universitat Pompeu Fabra, Spain

<https://orcid.org/0000-0003-0923-9192>

**Lluís Codina**

Universitat Pompeu Fabra, Spain

<https://orcid.org/0000-0001-7020-1631>

**Juan-José Boté-Vericad**

Universitat de Barcelona, Spain

<https://orcid.org/0000-0001-9815-6190>

---

Lopezosa, C., Aguilera-Cora, E., Codina, L., & Boté-Vericad, J. J. (2025). Web of Science Research Assistant: Functional analysis and usage recommendations. In J. Guallar, M. Váñez, & A. Ventura-Cisquella (Coords). *Digital communication. Trends and good practices* (pp. 174-189). Ediciones Profesionales de la Información. <https://doi.org/10.3145/cuvicecom.13.eng>

---

## Abstract

This chapter provides a comprehensive functional analysis of the Web of Science Research Assistant, a generative AI tool integrated into Clarivate's academic database. Designed to enhance research workflows, the tool supports tasks such as literature reviews, topic exploration, expert identification, and journal selection. Through a detailed examination of its interface and functionalities, including thematic summarization, co-citation mapping, and trend visualization, this chapter highlights the tool's potential to streamline scientific discovery. Several practical use cases are presented, demonstrating its capabilities in optimizing searches, identifying seminal papers, and suggesting relevant publication venues. Despite its innovative features, the tool's use must be guided by critical thinking, transparency, and ethical considerations. Researchers are encouraged to view the assistant not as a substitute, but as a complement to scholarly inquiry. The chapter concludes by emphasizing the value of reflective, layered interaction with AI to responsibly integrate these technologies into academic practice.

## Keywords

Generative artificial intelligence; Academic search tools; Web of Science; Web of Science Research Assistant; Literature review automation; Research ethics and transparency.

---

## 1. Introduction

Web of Science is one of the most important academic databases in the world, recognized for its multidisciplinary nature and the quality of its selection criteria. It is a database that covers everything from the experimental sciences to the humanities, including the social sciences. Web of Science covers documents published from 1900 to the present day in nearly 23,000 scientific journals, totalling around 180 million articles, as well as about 150,000 books (Clarivate, 2025a).

At the same time as databases like Web of Science have grown in importance, so too have generative artificial intelligence tools, which have rapidly expanded and now permeate all areas of society — especially the scientific and academic spheres. This chapter introduces Web of Science Research Assistant, the artificial intelligence integrated into the Web of Science database.

The following sections describe the tool's interface and analyze its functionalities. As will be shown, Web of Science Research Assistant has the potential to transform access to, exploration of, analysis of, and synthesis of scientific information. However, its use must always be accompanied by critical thinking and an ethical perspective, which includes transparency.

## 2. What Is Web of Science AI Research Assistant and how does It work?

Web of Science AI Research Assistant is a tool developed by Clarivate that uses generative AI models trained on the Web of Science Core Collection database to support researchers and students seeking to advance their academic work (Clarivate, 2025b).

This service goes beyond a traditional academic database, functioning instead as an assistant that can help (1) find seminal and relevant articles on a discipline or topic in seconds, (2) streamline advanced tasks such as literature reviews, expert identification, or journal selection for publication, (3) explore connections between academic concepts, authors, and scientific articles through interactive visualizations like trend maps and co-citation networks, and even (4) perform scientific document searches using natural language and multiple languages, thus facilitating access to scientific information without the need for advanced search operators.

When a user submits a query, the tool retrieves the most relevant documents using semantic similarity algorithms and keyword searches. It then organizes the results based on relevance and generates responses or summaries using the content of the selected articles. Web of Science Research Assistant also allows both document-based searches and synthesis-type questions, tailoring the response format to the user's needs.

Like other tools such as Scopus AI (Aguilera-Cora et al., 2024a; 2024b), Elicit (Arroyo-Machado, 2024), Epsilon (Ren et al., 2025), Perplexity (Torres-Salinas & Arroyo-Machado, 2025), ChatGPT (Torres-Salinas et al., 2024; Boté-Vericad et al., 2024), Copilot (Lopezosa, 2023a; Boté-Vericad, 2024), or Scite (Codina, 2024), Web of Science Research Assistant reduces the manual effort required to construct complex searches and facilitates the identification of trends, research gaps, and non-obvious connections between scientific works.

In the following section, we present a functional analysis of the tool to illustrate, with examples, the possibilities offered by Web of Science Research Assistant for research.

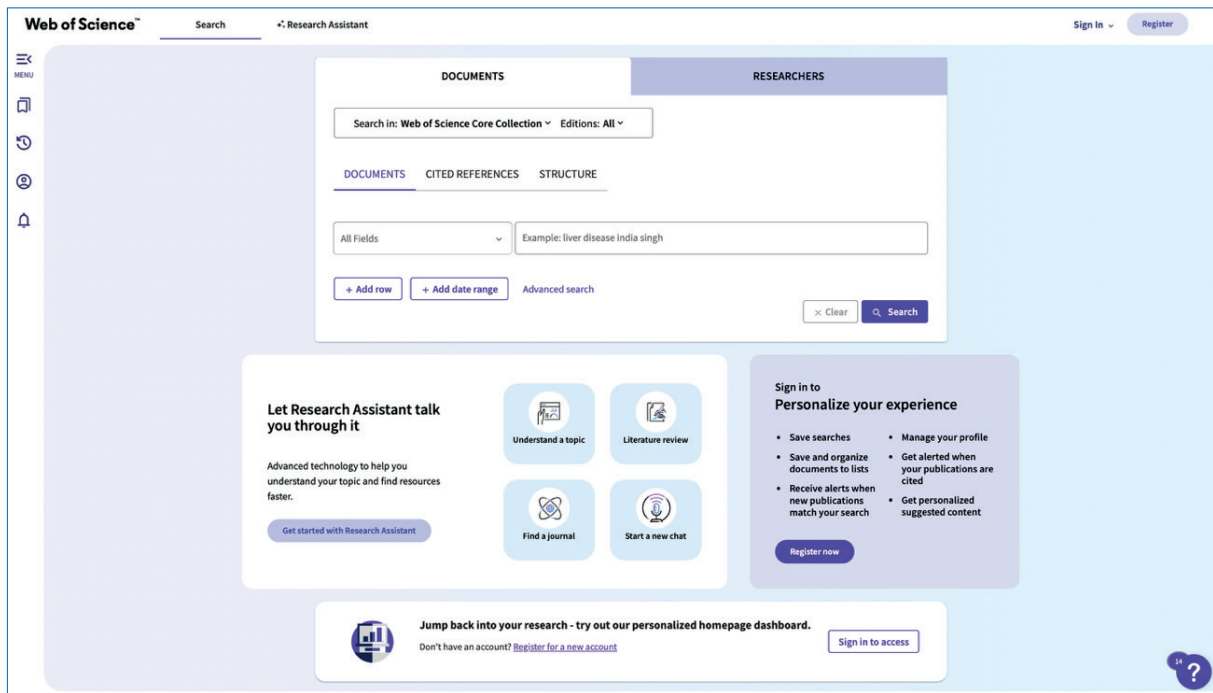
### **3. Practical use cases of Web of Science Research Assistant**

The Web of Science AI Research Assistant offers a variety of reports and visualizations designed to support academic research. These features allow users to efficiently analyze information and gain visual insights into relevant data and results.

First, when accessing the main Web of Science platform (Figure 1), the user is presented with the primary interface, focused on its academic search function. At the top, there is a menu that provides access to features such as "Search" and "Research Assistant," along with options to sign in or register. In our case, to activate the Web of Science AI service, we must select the "Research Assistant" tab.

**Figure 1**

Main Web of Science search interface with advanced options and research assistance.



Once we have accessed the “Research Assistant” resource, a new interface appears in the form of an interactive chat (Figure 2). In the centre of the screen, the virtual assistant initiates the conversation with the message: “I am going to walk you through understanding a topic. What is your topic of interest or research question?,” inviting the user to enter their topic of interest or research question.

**Figure 2**

Web of Science Research Assistant interface, guiding the user in understanding a topic based on an initial question.



On the left, there is a sidebar with the menu and the "Chat history" section, where the current conversation appears under the title "Understand a topic." Navigation icons and system options are also visible.

In the lower-left corner, there are links to send feedback, take a tour, or access information about the research assistant.

Once the "Research Assistant" interface has been described, we test it with a prompt: "What is the relationship between SEO and the digital media?," indicating that the topic to explore is the relationship between search engine optimization (SEO) and digital media.

The result of this prompt (Figure 3) shows the response generated by the tool. At the top, there is a section titled "Overview," where the main topics identified in the abstracts of the consulted articles are explained. In our case, these topics include search engine optimization (SEO) techniques, content and metadata features, and the impact of search engines on digital media. The overview also highlights the importance of SEO for improving web visibility, the role of content in search rankings, and how these dynamics affect journalism and academic research.

**Figure 3**

*Thematic summary generated by the Web of Science Assistant on the relationship between SEO and digital media, accompanied by key articles on optimization techniques and web visibility.*

The screenshot displays the Web of Science Research Assistant interface. At the top, a search bar contains the query "What is the relationship between SEO and the digital media?". Below the search bar, the "Overview" section provides a summary of the main themes identified in the provided abstracts: "search engine optimization techniques", "content characteristics and metadata", and "impact of search engines on digital media". It highlights strategies for improving web visibility, the role of content and metadata in search rankings, and the influence of search engines on journalism and academic research.

Below the overview, the "Search engine optimization techniques:" section lists several key points:

- Various techniques, such as PageRank and the use of in-links, are crucial for achieving high rankings in search engines like Google, as demonstrated in a study of optimized web pages <sup>92</sup>.
- Successful search engine optimizers recognize the importance of directories and social bookmarking sites in enhancing page visibility <sup>92</sup>.
- Professional communicators are encouraged to analyze keywo... [See full response](#)

Below the list, there are icons for "Copy Text" and "Here are some seminal papers:". The list of seminal papers includes:

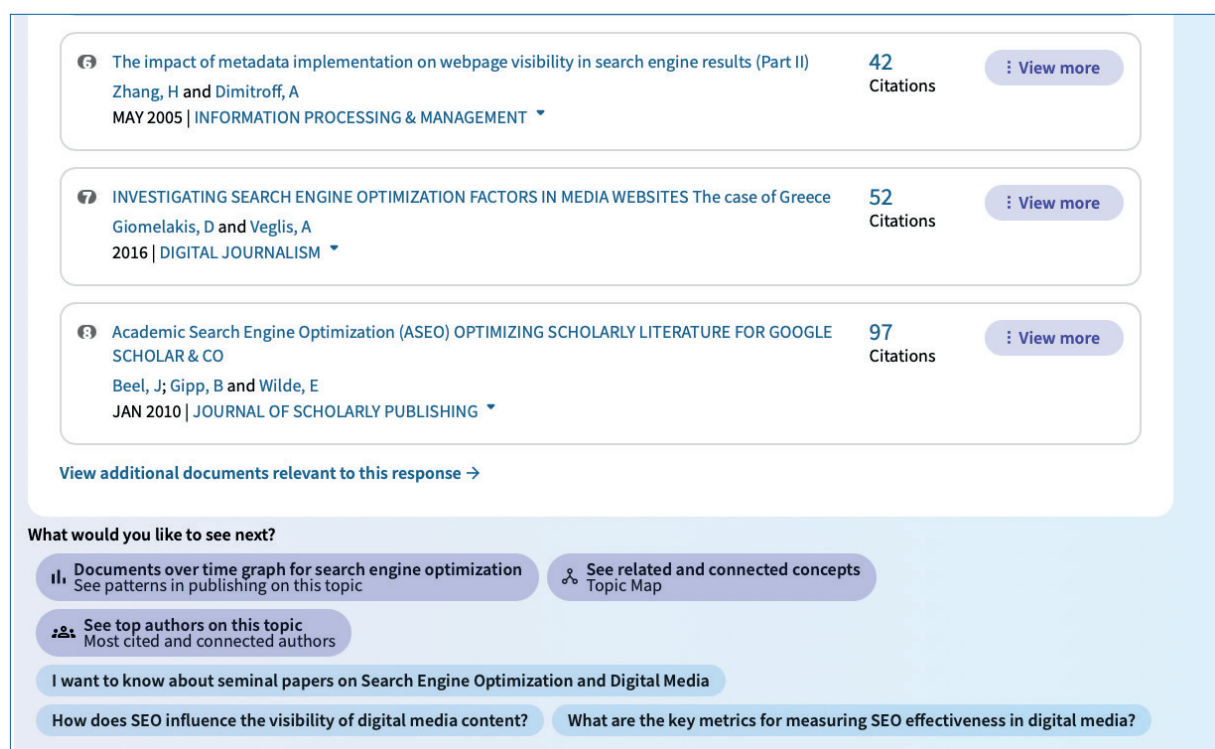
Paper Title	Author(s)	Year	Journal	Citations	View more
The anatomy of a large-scale hypertextual Web search engine	Brin, S and Page, L	APR 1998	COMPUTER NETWORKS AND ISDN SYSTEMS	6,942	<a href="#">View more</a>
Analysing Google rankings through search engine optimization data	Evans, MP	2007	INTERNET RESEARCH	92	<a href="#">View more</a>
The impact of webpage content characteristics on webpage visibility in search engine results (Part 1)	Zhang, J and Dimitroff, A	MAY 2005	INFORMATION PROCESSING & MANAGEMENT	54	<a href="#">View more</a>
How to Use Search Engine Optimization Techniques to Increase Website Visibility	Killoran, JB	MAR 2013	IEEE TRANSACTIONS ON PROFESSIONAL COMMUNICATION	57	<a href="#">View more</a>
What is search engine optimization: SEO?	Yalçın, N and Köse, U	2010	WORLD CONFERENCE ON LEARNING, TEACHING AND ADMINISTRATION PAPERS	47	<a href="#">View more</a>

Below the summary, a list of fundamental articles (“Seminal papers”) related to the topic is provided. Among them are key works such as Brin and Page (1998), with more than 6,900 citations, and other studies on Google rankings, content characteristics, and techniques to increase online visibility (Figures 3 and 4).

Each entry displays the article title, authors, source, publication date, number of citations, and a button to view more details. Additionally, below this list, there is a button to view more documents and an interactive section titled “What would you like to see next?,” offering complementary exploration options. These include analyzing the evolution of publications over time, viewing related concepts through a topic map, checking the most cited authors, and accessing new specific prompts that provide a guided and exploratory approach to the assistant.

**Figure 4**

*Second part of the thematic summary generated by the Web of Science Assistant on the relationship between SEO and digital media.*



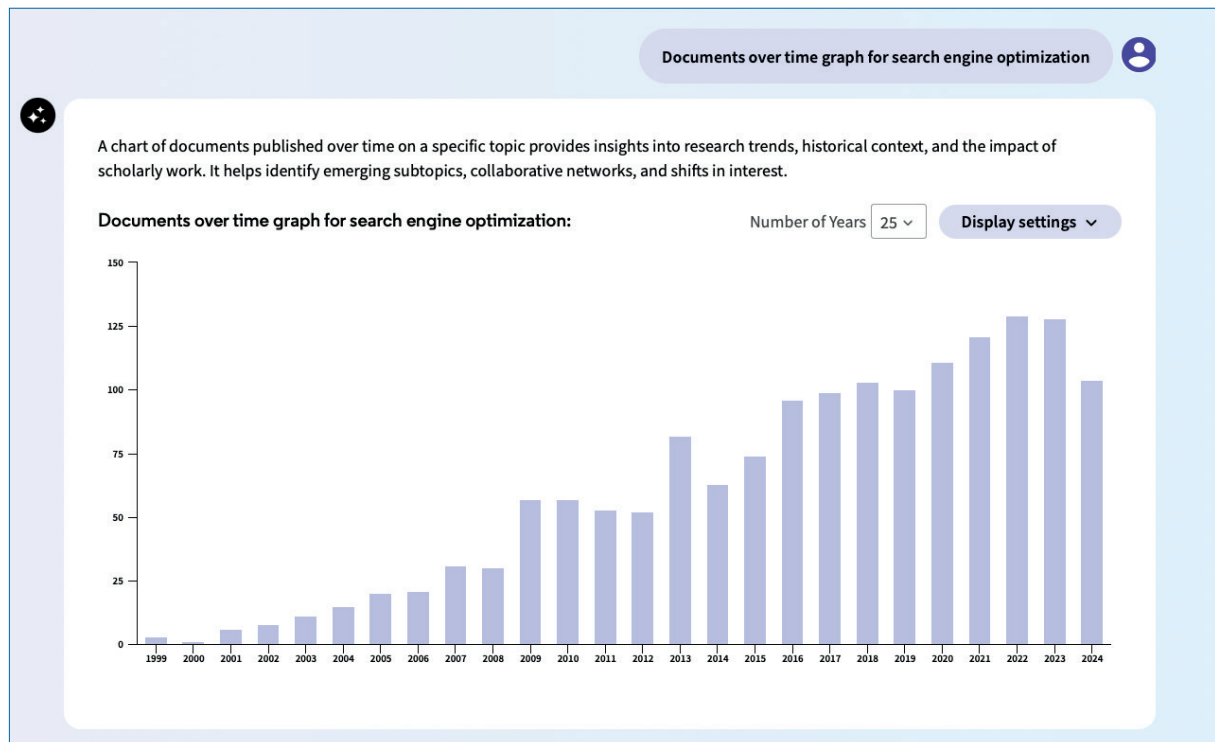
In addition to these options, the Research Assistant suggests other prompts such as “Documents over time graph for search engine optimization” (Figure 5). This feature displays a graphical visualization. At the top, Clarivate’s chatbot notes that this type of chart allows users to observe publication trends on a specific topic over time, offering historical context, identifying emerging subtopics, and revealing shifts in academic interest. Just beside the explanation, users can adjust the chart parameters using a dropdown menu for the number of years (in this case, 25) and another for visualization settings.

In the example shown, the bar chart illustrates the evolution of the number of documents published on “search engine optimization” from 1999 to 2024. A steady increase is observed starting in 2000, with a notable rise beginning around 2008. The peak in publications occurs in 2022 and 2023, with over 125 documents per year, with a slight decline in 2024. This type of analysis helps identify key moments of heightened research interest in SEO, possibly linked to technological advancements, changes in search engine algorithms, or the growth of digital communication.



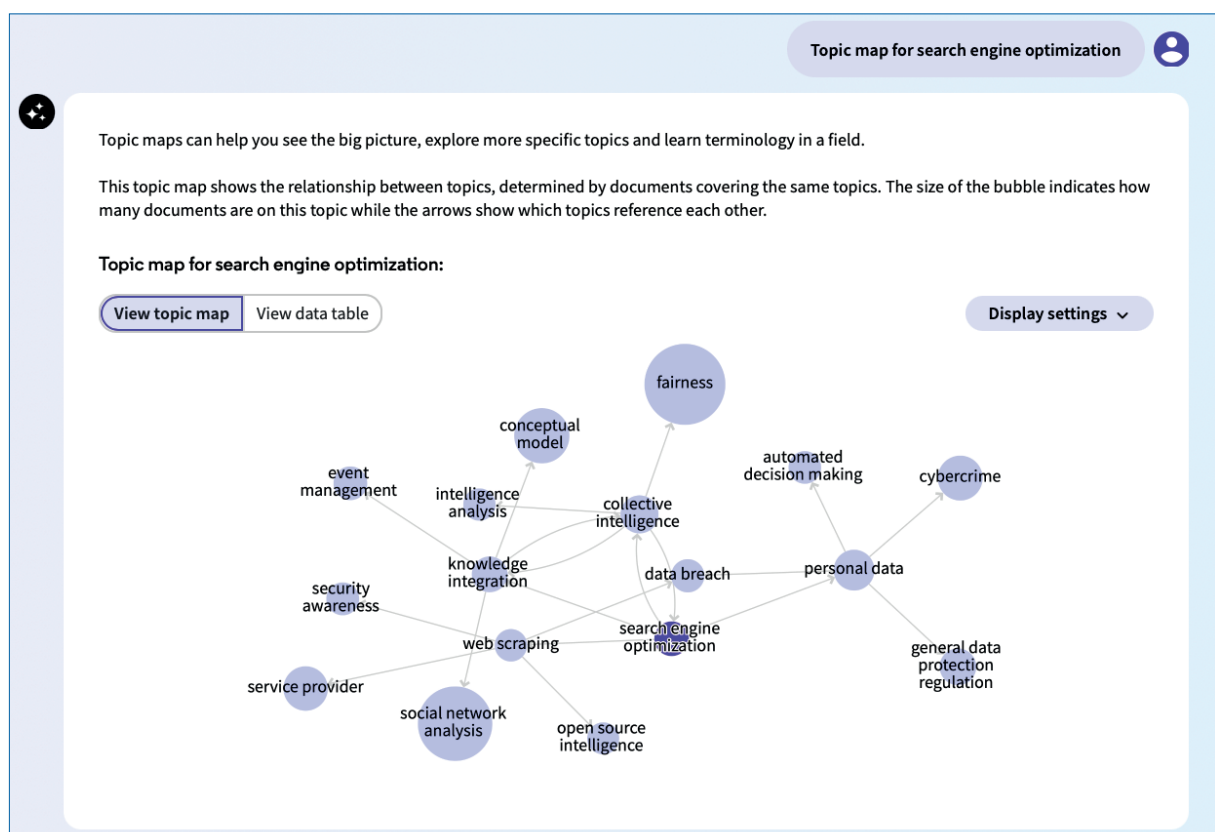
**Figure 5**

Graph of publications on search engine optimization (SEO) over time, showing the sustained growth of academic interest from 1999 to 2024.



**Figure 6**

Topic map on SEO visualizing related concepts and their interconnection, based on the co-occurrence of themes in the analyzed documents.

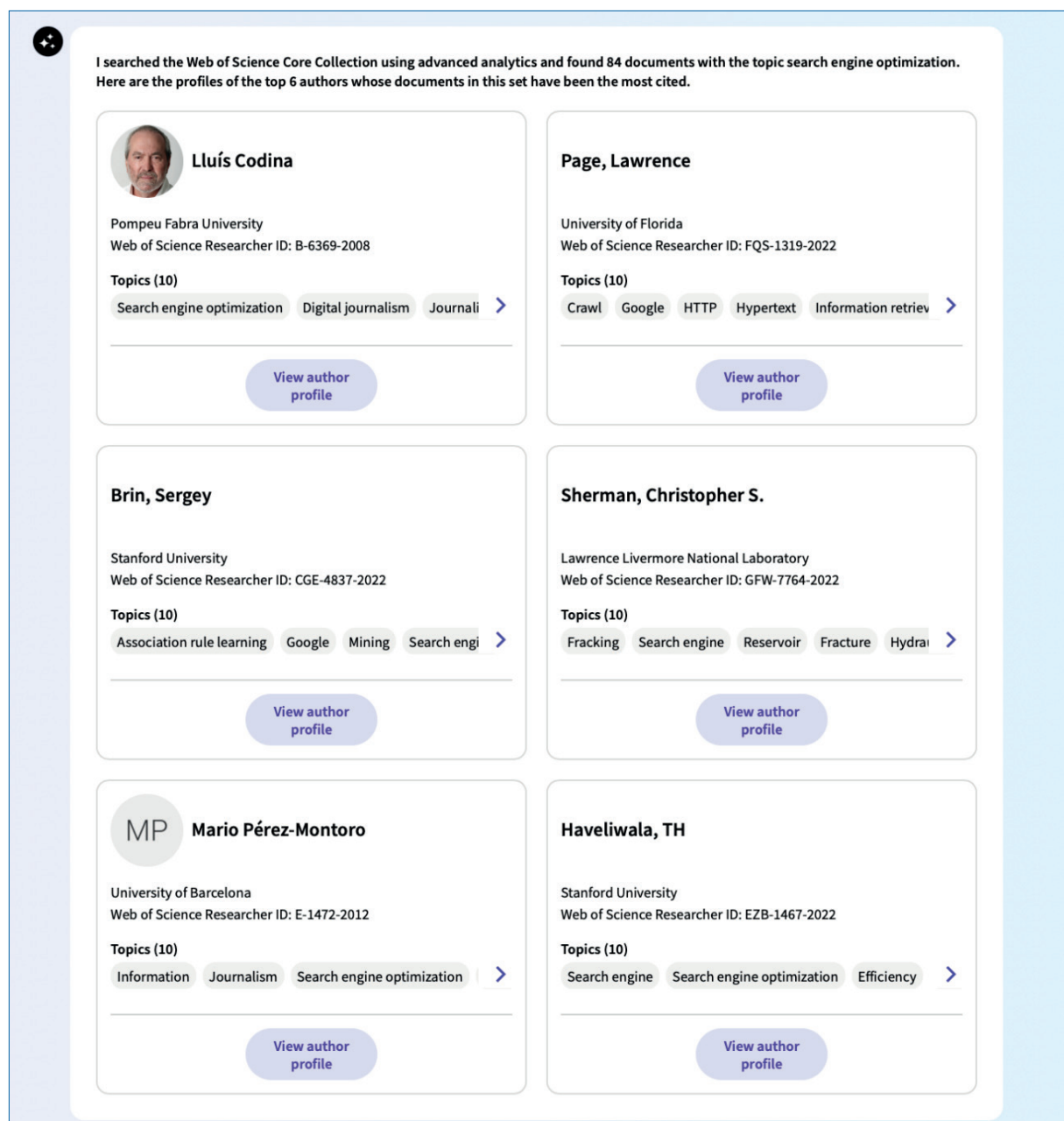


Additionally, we can create an interactive topic map. In the example presented, we once again use the topic of search engine optimization (Figure 6). At the top, it is explained that this type of visualization allows users to see the relationship between concepts within a research field, based on the documents covering those topics. The bubbles represent different themes, and their size indicates the number of documents related to each one, while the arrows show how the concepts reference one another.

In our example, the central node of the map is search engine optimization, which is connected to terms such as data breach, collective intelligence, web scraping, knowledge integration, open source intelligence, and personal data. Other related and relevant topics are also displayed, such as cybercrime, general data protection regulation, automated decision making, fairness, social network analysis, security awareness, and intelligence analysis.

**Figure 7**

*Profiles of the six most cited authors in Web of Science related to the topic of search engine optimization (SEO).*



This feature is particularly valuable, as it offers a form of representation that facilitates the exploration of subtopics, the identification of interdisciplinary connections, and the discovery of emerging trends in the field of SEO.

Another valuable feature offered by the Research Assistant is the most cited authors panel (Figure 7). In our example, for the topic of search engine optimization, 84 relevant documents were identified, and the profiles of the six authors whose works have been most cited within that set are displayed. Each author is represented in an individual card that includes their name, institutional affiliation, Web of Science researcher ID, the most frequent topics in their publications (up to 10), and a button to view their full profile. This panel provides a clear overview of the leading researchers in the field and allows users to explore their contributions in greater depth.

**Figure 8**

*Thematic and bibliographic analysis on SEO and digital media, including a summary of key techniques, impact on journalism, and references to seminal articles.*

The screenshot displays the Web of Science Research Assistant interface. At the top, a search bar contains the query: "I want to know about seminal papers on Search Engine Optimization and Digital Media". Below the search bar, the "Overview" section provides a summary of the main themes identified in the provided abstracts: "search engine optimization techniques", "content characteristics and metadata", and "impact of search engines on digital media". These themes highlight strategies for improving web visibility, the role of content and metadata in search rankings, and the influence of search engines on journalism and media practices.

The "Search engine optimization techniques:" section lists three key points:

- Various techniques, such as PageRank and the use of in-links, are crucial for achieving high rankings in search engines like Google, as demonstrated in a study of optimized web pages.
- Successful search engine optimizers recognize the importance of directories and social bookmarking sites in enhancing page visibility.
- Continuous adaptation to changing algorithms and practices is essential for effective search engine optimization.

The "Content characteristics and metadata:" section lists three key points:

- Key content characteristics, including keyword placement and frequency, significantly impact webpage visibility in search engine results.
- The implementation of metadata, particularly the subject field, plays a vital role in improving a webpage's visibility, with effective keywords derived from both the title and full text.
- A combination of content strategies, including keyword analysis and engagement with other web content creators, can enhance search engine rankings.

The "Impact of search engines on digital media:" section lists three key points:

- The evolution of digital technologies has transformed journalism, making search engine visibility critical for driving traffic to news websites.
- Media professionals are increasingly utilizing SEO practices to enhance the reach and effectiveness of their content.
- The concept of academic search engine optimization (ASEO) has emerged, focusing on optimizing scholarly literature for better visibility in academic search engines.

The "Conclusion" section summarizes the findings, stating that the abstracts collectively emphasize the significance of "search engine optimization techniques", "content characteristics and metadata", and "impact of search engines on digital media". These themes illustrate the strategies and considerations necessary for improving web visibility and the evolving role of search engines in various fields, particularly in journalism and academia.

Below the conclusion, there are options to "See less" and "Copy Text".

The "Here are some seminal papers:" section displays a list of three seminal papers:

Paper Title	Author(s)	Date	Journal	Citations	Action
The anatomy of a large-scale hypertextual Web search engine	Brin, S and Page, L	APR 1998	COMPUTER NETWORKS AND ISDN SYSTEMS	6,942	View more
Analysing Google rankings through search engine optimization data	Evans, MP	2007	INTERNET RESEARCH	92	View more
The impact of webpage content characteristics on webpage visibility in search engine results (Part 1)	Zhang, J and Dimitroff, A	MAY 2005	INFORMATION PROCESSING & MANAGEMENT	54	View more

Another very useful instruction that can be given to the Web of Science AI Assistant is to identify seminal papers on specific topics. In our case (Figure 8), we used the following prompt: "I want to know about seminal papers on Search Engine Optimization and digital media."

The result is a section titled "Overview," which summarizes three key themes extracted from the abstracts of the analyzed documents: (1) search engine optimization techniques, (2) content and metadata characteristics, and (3) the impact of search engines on digital media. The text provided by the Web of Science AI Research Assistant highlights the importance of techniques such as PageRank, metadata optimization, collaboration among content creators, and ongoing adaptation to algorithmic changes to enhance web visibility.

The result of this prompt also analyzes how SEO has influenced journalism and digital media, emphasizing the increasing use of these strategies by professionals to expand the reach of their content. In addition, it mentions the emergence of the concept of ASEO (Academic Search Engine Optimization) to improve the visibility of academic publications. The response concludes with a summary that reinforces the relevance of these three core areas in the fields of journalism and academic research. Finally, it lists seminal papers, including their title, authors, source, year, number of citations, and a button to access more information (Figures 8 and 9).

**Figure 9**

*Continuation of the thematic and bibliographic analysis on SEO and digital media, including a summary of key techniques, impact on journalism, and references to seminal articles.*

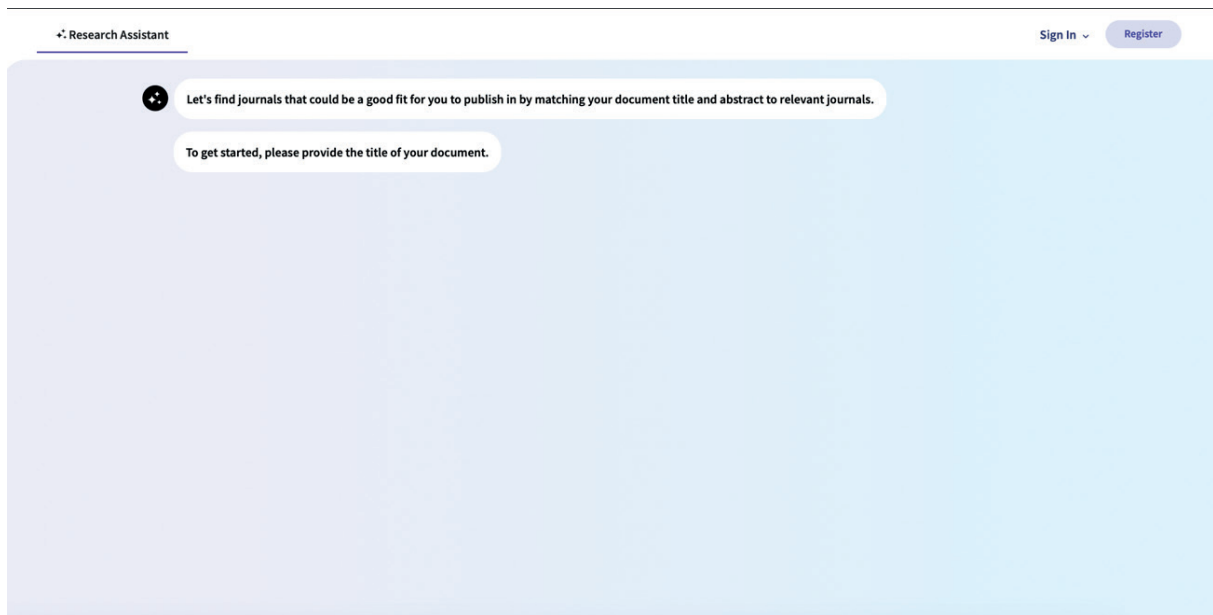
4	How to Use Search Engine Optimization Techniques to Increase Website Visibility Killoran, JB MAR 2013   IEEE TRANSACTIONS ON PROFESSIONAL COMMUNICATION	57 Citations	<a href="#">View more</a>
5	What is search engine optimization: SEO? Yalçın, N and Köse, U 2010   WORLD CONFERENCE ON LEARNING, TEACHING AND ADMINISTRATION PAPERS	47 Citations	<a href="#">View more</a>
6	The impact of metadata implementation on webpage visibility in search engine results (Part II) Zhang, H and Dimitroff, A MAY 2005   INFORMATION PROCESSING & MANAGEMENT	42 Citations	<a href="#">View more</a>
7	INVESTIGATING SEARCH ENGINE OPTIMIZATION FACTORS IN MEDIA WEBSITES The case of Greece Giromelakis, D and Veglis, A 2016   DIGITAL JOURNALISM	52 Citations	<a href="#">View more</a>
8	Academic Search Engine Optimization (ASEO) OPTIMIZING SCHOLARLY LITERATURE FOR GOOGLE SCHOLAR & CO Beel, J; Gipp, B and Wilde, E JAN 2010   JOURNAL OF SCHOLARLY PUBLISHING	97 Citations	<a href="#">View more</a>

[View additional documents relevant to this response →](#)

On the other hand, Web of Science Research Assistant includes a feature that suggests suitable academic journals for publishing a specific scholarly work (Figure 10). When accessing this resource, a message from the assistant appears at the centre of the screen, stating: "Let's find journals that could be a good fit for you to publish in by matching your document title and abstract to relevant journals." The user is then prompted to provide the title of their document to begin the process.

**Figure 10**

*Web of Science Assistant for journal suggestions, recommending suitable publications based on the document's title and abstract.*



In our case, we entered the title of a manuscript: "Analysis of Google News coverage: A comparative study of Brazil, Colombia, Mexico, Portugal, and Spain" (Figure 11). In response, the assistant asks for the document's abstract, indicating that it should be a brief description of the content and research, approximately 100 words (Figure 12).

At the bottom, there is a text input field where the user must type the abstract, accompanied by an arrow button to submit it.

**Figure 11**

*Step in the Web of Science Assistant journal suggestion process, where the abstract is requested after entering the study title.*

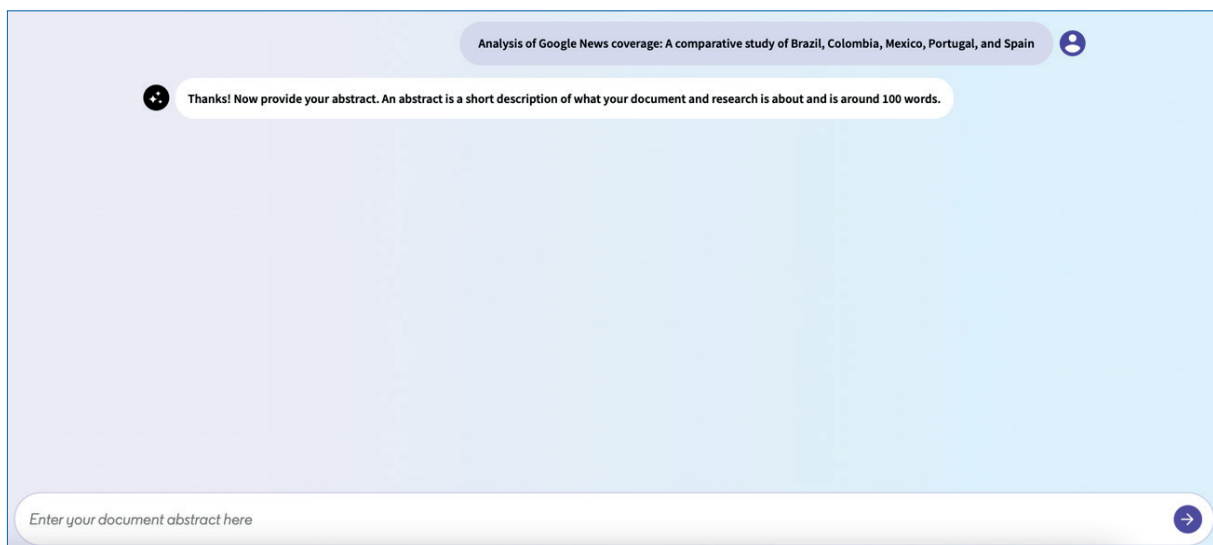




Figure 12

Example of an abstract entered into the Web of Science Assistant to receive journal recommendations, focused on a comparative analysis of news coverage in Google News across five Ibero-American countries.

+

Thanks! Now provide your abstract. An abstract is a short description of what your document and research is about and is around 100 words.

This study aims to examine the news coverage provided by Google News across five Ibero-American countries, including three from Latin America and two from Europe: Brazil, Colombia, Mexico, Portugal, and Spain. The main focus is to highlight the differences and similarities in news presentation within diverse contexts, evaluating the presence and distribution of news using quantitative indicators, and analyzing the predominant content in each country's news, based on a dataset collected between January 2 and January 31, 2024. This includes examining news sources, geographic coverage, prominent figures, and the prevalence of sensational elements through the identification of clickbaits. Our research employs statistical analyses and algorithmic solutions in natural language processing and artificial intelligence to generate results. The analyses revealed consistency in the daily delivery of content by Google News, with specific variations in the update rate across the studied countries. A diversity of news sources was observed, with a greater tendency toward local news and frequent mentions of politicians, celebrities, and businesspeople. In addition, the analyses uncovered a significant presence of clickbait, with variations across the countries and topics.

u

Figure 13

Results from the Web of Science Assistant showing the five most compatible journals for publication, based on the analyzed document's title and abstract.

+

Thank you! I'm now searching for journals that match your title and abstract based on keywords.

+

Here are the top 5 journals that match your documents.

**SUSTAINABILITY**

Match Score 0.50

Publisher

MDPI,ST ALBAN-ANLAGE 66, BASEL, SWITZERLAND, CH-4052

ISSN / eISSN

2071-1050 / 2071-1050

Web of Science Core Collection

Science Citation Index Expanded | Social Sciences Citation Index

Additional Indexes

Current Contents Agriculture, Biology & Environmental Sciences | Current Contents Social And Behavioral Sciences | Essential Science Indicators

**ENVIRONMENTAL SCIENCE AND POLLUTION RESEARCH**

Match Score 0.25

Publisher

SPRINGER HEIDELBERG,TIERGARTENSTRASSE 17, HEIDELBERG, GERMANY, D-69121

ISSN / eISSN

0944-1344 / 1614-7499

Additional Indexes

Biological Abstracts | BIOSIS Previews

**FRONTIERS IN PSYCHOLOGY**

Match Score 0.22

Publisher

FRONTIERS MEDIA SA,AVENUE DU TRIBUNAL FEDERAL 34, LAUSANNE, SWITZERLAND, CH-1015

ISSN / eISSN

1664-1078 / 1664-1078

Web of Science Core Collection

Social Sciences Citation Index

Additional Indexes

Current Contents Clinical Medicine | Current Contents Social And Behavioral Sciences | Essential Science Indicators

**INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS**

Match Score 0.19

Publisher

SCIENCE & INFORMATION SAI ORGANIZATION LTD,19 BOLLING RD, BRADFORD, WEST YORKSHIRE, ENGLAND, 00000

ISSN / eISSN

2158-107X / 2156-5570

Web of Science Core Collection

Emerging Sources Citation Index

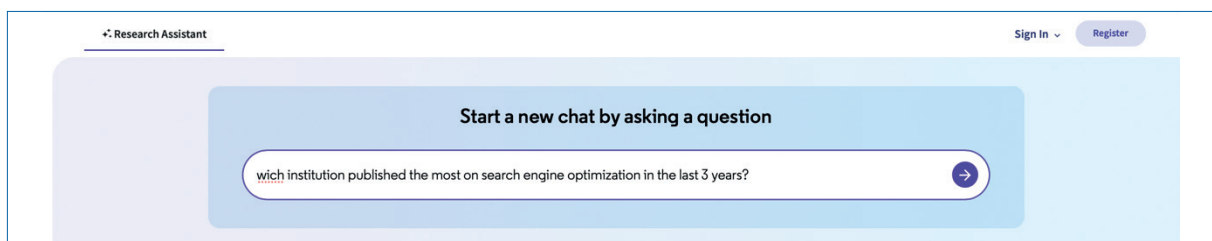
As a result of providing the title and abstract, the Web of Science AI tool presents a list of the five scientific journals that best match the content of the document, along with their respective match scores, which indicate the level of thematic alignment based on keywords and indexing areas (Figure 13). This is the last step in the journal recommendation process, offering the user publication options based on thematic relevance, academic visibility, and disciplinary coverage.

Another interesting instruction that can be given to the Research Assistant is to find out which institutions have conducted the most research on a specific topic (Figure 14). To do this, we can enter the following prompt: "Which institution has published the most on search engine optimization in the last 3 years?"

In this case, we aim to determine which institution has published the most on SEO over the past three years. This function enables the assistant to retrieve and analyze recent publications in order to identify the leading institutions in research on that specific topic.

**Figure 14**

*Initial query in the Web of Science Assistant to identify which institution has published the most on SEO in the past three years.*



The result (Figure 15) is a bar chart. At the top, it is explained that the search was conducted using publications from November 29, 2021, to November 29, 2024, based on a combination of terms related to search engine optimization. It is also noted that the query returned 137,295 documents.

The chart, titled "Institution bar chart," displays the 10 institutions with the highest number of publications on the topic. In our example, the Egyptian Knowledge Bank (EKB) stands out significantly, with over 5,500 publications, occupying the top position. This visualization offers a quick understanding of which institutions lead scientific production on SEO and related topics during the specified period. In addition to the bar chart, this instruction can also be viewed in "tree map" mode (Figure 16).



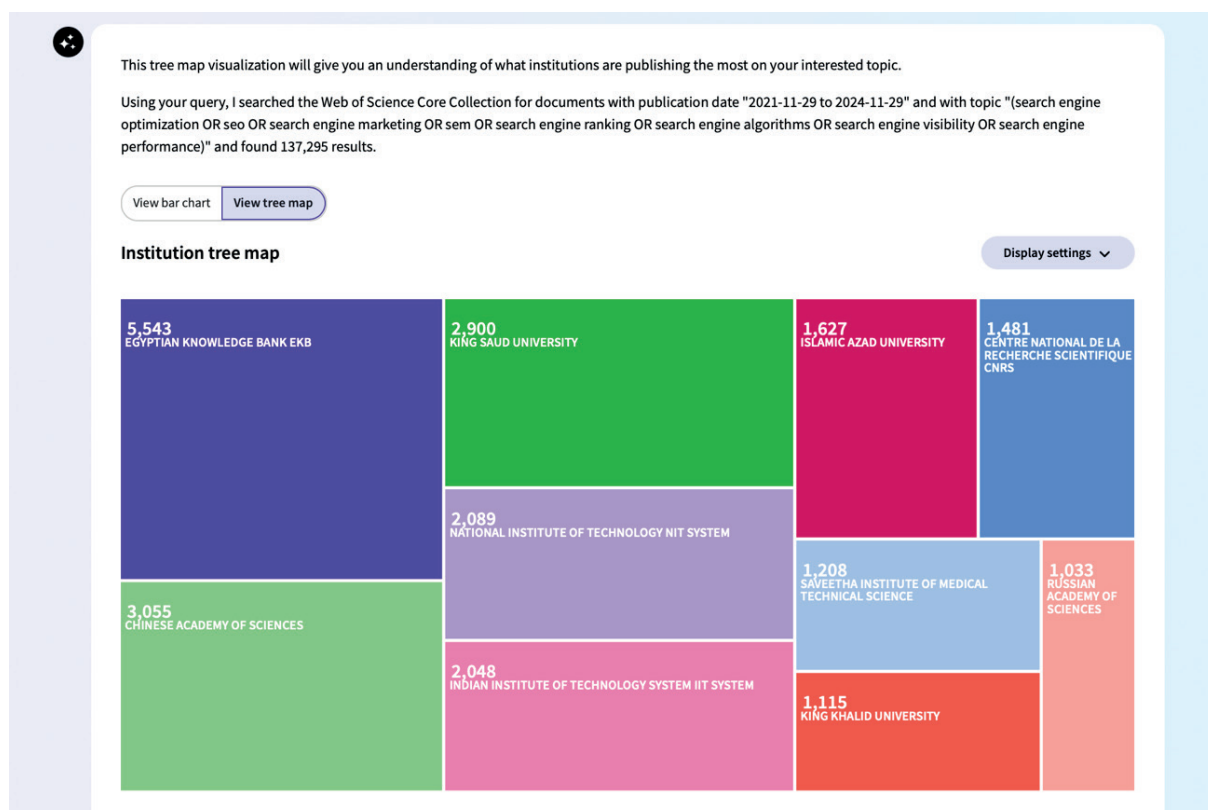
**Figure 15**

Bar chart showing the institutions that have published the most on SEO in the past three years, highlighting the Egyptian Knowledge Bank as the most productive.



**Figure 16**

Tree map showing the institutions with the highest volume of publications on SEO between 2021 and 2024, led by the Egyptian Knowledge Bank.



Throughout this section, we have presented several examples of what can be done with Clarivate's AI Assistant for the Web of Science database. Although the current capabilities are already extensive, it is likely that the features offered by this tool will continue to improve and expand in the future, as AI technology continues to evolve.

## 4. Conclusions

As we have seen throughout this section, Web of Science Research Assistant offers a comprehensive and integrated experience that combines article summaries and analysis, personalized reports, interactive visualizations, source management and citation, and real-time feedback. All of this makes it a very interesting complement for conducting research.

It is an advanced solution that could ultimately transform the way researchers access, explore, and manage scientific information. However, although tools like this can be very useful, they do not replace the researcher's critical work. The selection of sources, the deep analysis of texts, and the construction of a solid theoretical framework remain human tasks.

In addition, it is important to take into account aspects such as ethics, transparency, and critical thinking when using artificial intelligence in research (Lopezosa, 2023b; Codina, 2025). Not everything the tool suggests is always valid, nor should all the information it offers be accepted without review (Orduña-Malea & Cabezas-Clavijo, 2023; Font-Julià et al., 2024). In short, the existence of the Web of Science Assistant should be understood as very good news, because it helps researchers do their work better, with less repetitive effort and in less time. Therefore, it is an excellent tool for starting or advancing a project.

The key, as always when using AI, is threefold: (1) not to delegate all the work to AI, but to use it to expand and guide, without allowing it to replace the researcher's role; (2) to verify assertions, sources, and factual or conceptual data; and (3) to be transparent about its use. Used in this way — as a process of layered, critical interaction — it can become a tool that permanently transforms certain ways of doing science.

## 5. Funding

This work is part of the Project "Parameters and strategies to increase the relevance of media and digital communication in society: curation, visualisation and visibility (CUVICOM)". Grant PID2021-123579OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU.

## 6. References

Aguilera-Cora, E., Lopezosa, C., & Codina, L. (2024). *Scopus AI Beta Functional analysis and cases*. Barcelona: Universitat Pompeu Fabra. Departament de Comunicació. (Serie Editorial DigiDoc. DigiDoc Reports).

Aguilera-Cora, E., Lopezosa, C., Fernández-Cavia, J., & Codina, L. (2024). Accelerating research processes with Scopus AI: A place branding case study. *Revista Panamericana de Comunicación*, 6(1). <https://doi.org/10.21555/rpc.v6i1.3088>

Arroyo-Machado, W. (2024). Búsqueda y análisis bibliográfico con Elicit. *Infonomy*, 2(4). <https://doi.org/10.3145/infonomy.24.050>

Boté-Vericad, J.-J., Fabeiro, R., & Anglada, L. (2024). *Creando un chatbot con ChatGPT como soporte a la catalogación en bibliotecas, archivos y centros de documentación. Comparación de modelos de lenguaje en versión gratuita y premium*. <https://hdl.handle.net/2445/213600>

Boté-Vericad, J.-J. (2024). *Copilot Microsoft y ChatGPT: Análisis de datos, dispersión, relación y correlaciones de variables* [vídeo]. <https://hdl.handle.net/2445/213581>

Clarivate. (2025a). *Resources for Librarians and Administrators*. <https://clarivate.libguides.com/librarianresources/coverage>

Clarivate (2025b). *Web of Science Research Assistant*. <https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/web-of-science-research-assistant>

Codina, L. (2024). Inteligencia artificial generative para trabajos académicos: método de análisis y comparación de Elicit, Epsilon, Perplexity, SciSpace y Scite. *Lluís Codina*. <https://www.lluiscodina.com/inteligencia-artificial-generativa-academia>

Codina, L. (2025). Interacción crítica escalonada con inteligencia artificial: explorando caminos para usar la IA en la academia. *Lluís Codina*. <https://www.lluiscodina.com/interaccion-critica-ia>

Font-Julián, C. I., Orduña-Malea, E., & Codina, L. (2024). ChatGPT Search as a tool for scholarly tasks: evolution or devolution?. *Infonomy*, 2(5). <https://doi.org/10.3145/infonomy.24.059>

Lopezosa, C. (2023a). Bing chat: hacia una nueva forma de entender las búsquedas. *Anuario ThinkEPI*, 17. <https://doi.org/10.3145/thinkepi.2023.e17a04>

Lopezosa, C. (2023b). La Inteligencia artificial generativa en la comunicación científica: retos y oportunidades. *Revista de Investigación e Innovación en Ciencias de la Salud*, 5(1), 1-5. <https://doi.org/10.46634/riics.211>

Orduña-Malea, E., & Cabezas-Clavijo, Á. (2023). ChatGPT and the potential growing of ghost bibliographic references. *Scientometrics*, 128(9), 5351-5355. <https://doi.org/10.1007/s11192-023-04804-4>

Ren, Y., Mohammadi, L., & Codina, L. (2025). Aplicación de sistemas de IA en un trabajo de final de máster universitario: presentación de un caso. Barcelona: Universitat Pompeu Fabra. Departamento de Comunicación. <http://hdl.handle.net/10230/70069>

Torres-Salinas, D., & Arroyo Machado, W. (2025). *Flujos de trabajo inteligentes con ChatGPT, Perplexity y NotebookLM*. <https://doi.org/10.5281/zenodo.14981646>

Torres-Salinas, D., Thelwall, M., & Arroyo Machado, W. (2024). ChatGPT for Bibliometrics: Potential applications and limitations. <https://hdl.handle.net/10481/92547>

The background of the entire page is a complex network of thin, intersecting lines in shades of blue, yellow, and red, forming a web-like structure. The top half of the page is white, and the bottom half is a solid blue color, with the network pattern visible in both.

# CUVICOM

Curación, Visualización y Visibilidad en la Comunicación Digital

# EPI